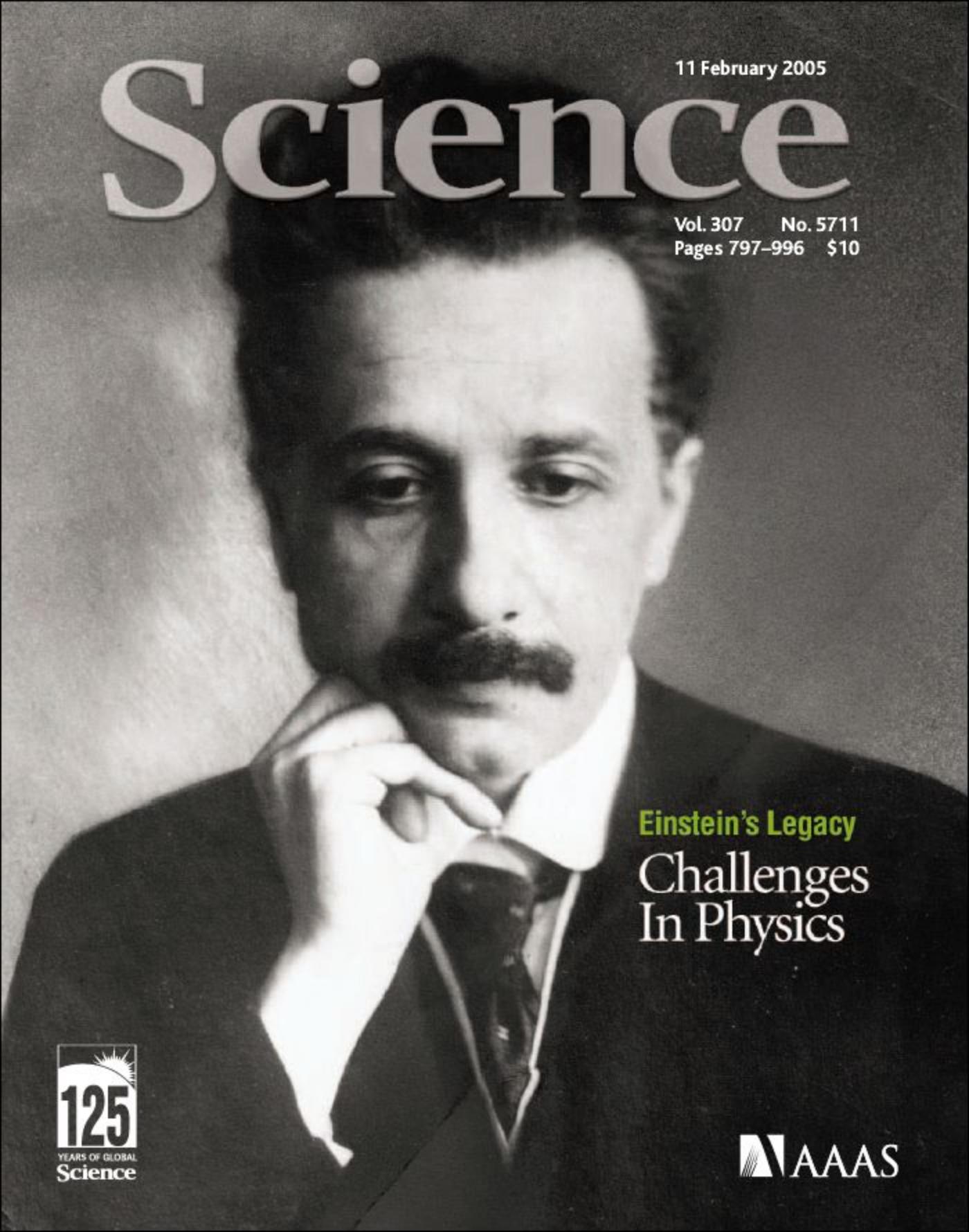
TEAM tAO

11 February 2005

Science

Vol. 307 No. 5711
Pages 797–996 $10



Einstein's Legacy

Challenges In Physics

125 YEARS OF GLOBAL Science

AAAS

Science

Special Issue

Einstein's Legacy

Volume 307
11 February 2005
Number 5711

Albert Einstein early in his career (circa 1910). In 1905, Einstein published five seminal papers on the photoelectric effect, special relativity, and Brownian motion. We celebrate the 2005 World Year of Physics that honors Einstein's achievements with a special section focusing on some of the challenges that remain in physics. [Photo: Hulton Archive/Getty Images]

For related online content in Next Wave, see page 809 or go to www.sciencemag.org/sciext/einstein

Departments

News of the Week

827 & 920

832

News Focus

Letters

Books *et al.*

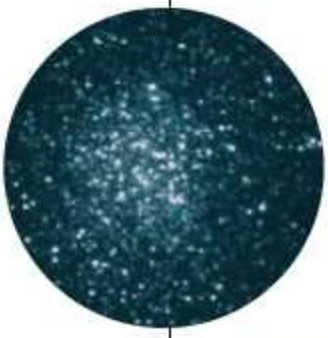
Contents continued

Science

Policy Forum

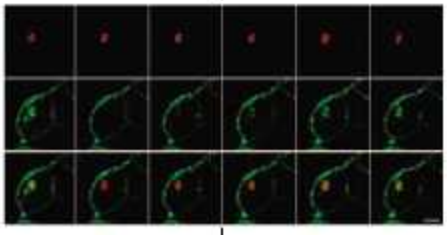
Perspectives

Essay



855 & 892

Science Express www.sciencexpress.org



Brevia

Research Articles

Reports

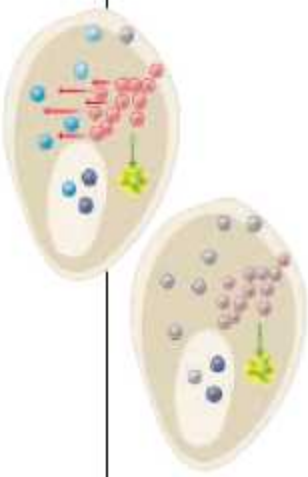
Young Scientist Award

863

Contents continued

REPORTS CONTINUED

861 & 910



858 & 925

AAAS
ADVANCING SCIENCE, SERVING SOCIETY

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals Mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices.

Domestic individual membership and subscription (51 issues): $135 ($74 allocated to subscription) Domestic institutional subscription (51 issues): $550; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery) $85. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request; GST #125488122. Publications Mail Agreement Number 1068628. Printed in the U.S.A.

Change of address: allow 4 weeks, giving old and new addresses and 8-digit account number. Postmaster: Send change of address to Science, P.O. Box 1811, Danbury, CT 06813–1811. Single copy sales: $10.00 per issue prepaid includes surface postage; bulk rates on request. Authorization to photocopy material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by AAAS to libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that $15.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. The identification code for Science is 0036-8075/03 $15.00. Science is indexed in the Reader's Guide to Periodical Literature and in several specialized indexes.

Contents continued

Science
www.scienceonline.org

sciencenow www.sciencenow.org Daily News Coverage

Solar System Lite
Objects not much larger than Jupiter may host planets of their own.

New Tuberculosis Strategy Shows Promise
Combination of drugs and DNA vaccine protects against a second infection.

The Sky Is Falling!
Close, but not quite, say astronomers.

Advanced physics at the beach.

science's next wave www.nextwave.org Career Resources for Young Scientists

Related Einstein's Legacy section page 865

- Global/Europe: **European Faces of Physics** *A. Forde*
  Going to the beach may help European physicists relate Brownian motion and temperature.
- Global/MiSciNet: **"We're Doing Just Fine"** *C. Parks*
  When people with different inputs don't participate in science; we may miss the opportunity to create jazz.
- Global/Canada: **Answering the Quantum Quest—Canada's Perimeter Institute** *A. Fazekas*
  A theoretical physics institute is gaining recognition in areas like quantum gravity and quantum computing.
- Global/US: **The Perimeter Institute's Quantum Mechanic** *J. Kling*
  Dr. Gottesman developed a mathematical language for dealing with errors made by quantum computers.

MiSciNet: **Being Prepared** *B. Gledhill*
Students at 2-year colleges get tips on making smooth transitions to 4-year institutions.

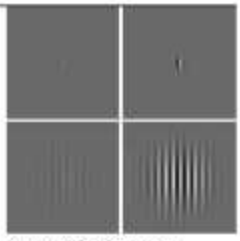
GrantsNet: **International Grants and Fellowships Index** *Edited by A. Kotok*
Get the latest list of international fellowship, research funding, and prize competition opportunities.

science's sage ke www.sageke.org Science of Aging Knowledge Environment

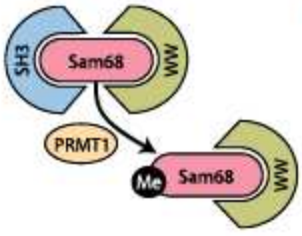
Perspective: **Drug Discovery in Neurodegenerative Diseases** *H. Geerts, J. Q. Trojanowski, V. M.-Y. Lee*
New research is meeting the need for drugs to counter neurodegeneration.

News Focus: **Dosed to Death** *R. J. Davenport*
Thyroid hormone treatments shorten mouse life span.

News Focus: **The Better to See You With** *M. Leslie*
Older folks surpass youngsters in discerning one kind of motion.



New directions for aging vision.



Methylation changes Sam68's binding partners.

science's stke www.stke.org Signal Transduction Knowledge Environment

Review: **Protein Interfaces in Signaling Regulated by Arginine Methylation** *F.-M. Boisvert, C. A. Chénard, S. Richard*
Covalent modification of arginine residues appears to regulate activity of disease-related gene products and various signaling proteins.

Teaching Resource: **Ras-MAPK Pathways** *A. Chan*
These lecture materials provide an introduction to signaling through kinase-mediated cascades.

Connections Map Overview: **Wnt/β-Catenin Pathway** *R. T. Moon*
Wnt pathways modulate cell proliferation, survival, behavior, and fate in both embryos and adults.

Connections Map Overview: ***Xenopus* Egg Wnt/β-Catenin Pathway** *R. T. Moon*
β-catenin signaling promotes polarization of the embryo to establish the dorsoventral axis.

Separate individual or institutional subscriptions to these products may be required for full-text access.

Stuffed with Pulsars

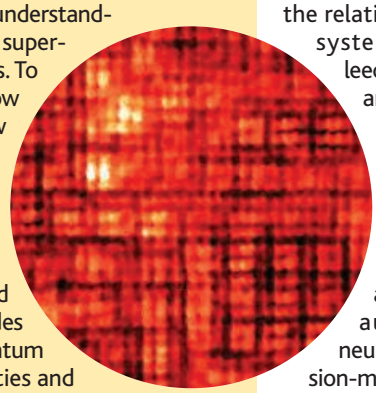
Globular clusters contain thousands to millions of stars and are among the oldest objects in the universe. **Ransom *et al.*** (p. 892; see the Perspective by **Lorimer**) studied the globular cluster, Terzan 5, with the Green Bank radio telescope and discovered 21 new millisecond pulsars, about half in binary systems (two with close enough orbits to allow repeated eclipses and others with unusually wide orbits or odd companions), several with some of the highest rates of rotation, and two with masses that exceed the theoretical limits for neutron stars. This menagerie of extraordinary pulsars has much to tell us about pulsar physics, general relativity, and globular cluster evolution.

Through a Glass, Darkly

Most carbon nanotubes are grown with the aid of catalyst particles that reside at the tips of the growing tubes. However, how do nanotubes grow during the catalyst-free process in which an arc is struck between two graphite rods? **De Heer *et al.*** (p. 907) studied this process in detail and found the formation of amorphous carbon beads on a small number of the multiwalled tubes, which suggests that the tubes grow in a manner similar to other crystal-growth processes.

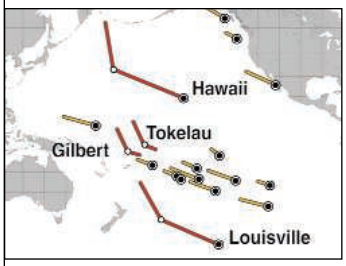
Cuprates in Real- and Momentum-Space

Recent real-space imaging experiments on the high-temperature cuprate superconductors have revealed the existence of a "checkerboard" charge-ordering pattern on the surface. This structure has received much attention in terms of its relation to understanding the mechanism underlying superconductivity in these materials. To strengthen the case, what is now needed are samples that allow direct comparison between real-space and momentum-space data. Working with the sodium-doped oxychloride superconductor, **Shen *et al.*** (p. 901) present angle-resolved photoemission data that provides complementary data in momentum space. Interpreting the similarities and differences found in the real-space and momentum-space experiments may provide some guidance in revealing the underlying mechanism.



Shifting Reference Frames

There are several chains of volcanoes and seamounts within the Pacific plate that have been used to track its motion, given the assumption that there is a fixed hot spot that can serve as a reference frame. The Hawaiian-Emperor chain shows a sharp bend which indicates that a change in growth direction occurred at about 47 million years ago. **Koppers and Staudigel** (p. 904) dated a similar bend in two volcanic chains in the southern Pacific plate and found that the chains changed directions at different times. The lack of synchronicity among the three bends in the three volcanic chains means that the hot spot must have been moving or the plate properties were different in different regions. These results indicate that a fixed hot-spot reference frame cannot be used to track plate motions and that some revisions of plate tectonic histories may be needed.



CREDITS (TOP TO BOTTOM): SHEN *ET AL.*; KOPPERS AND STAUDIGEL

Ear Origins

All living mammals have a distinctive ear containing three bones (hammer, anvil, and stirrup) and a single jaw bone. These structures evolved from four or more bones that made up the jaw of their reptilian ancestor in the Mesozoic. It has been thought that this evolution occurred in a basal mammal, prior to the split of monotremes (the few extant mammals that lay eggs) from marsupials and placentals. **Rich *et al.*** (p. 910; see the Perspective by **Martin and Luo**) now show that the ear of the earliest known monotreme, from the Early Cretaceous, has only one bone. Thus, the complex ears of mammals arose separately and converged in different mammalian lineages.

Decisions, Decisions...

What makes an individual decide to choose one set of activities over another? **Briggman *et al.*** (p. 896) tried to unravel the mechanisms underlying behavioral choice in the relatively simple nervous system of the medicinal leech. They presented an animal with a constant stimulus that repeatedly produced two different, mutually exclusive behaviors with roughly equal probabilities. This approach allowed the authors to focus on neurons involved in decision-making rather than the neural effects of sensory input, which was invariant. Neurons exhibiting decisive roles in the choice between swimming and crawling were identified by combining high-resolution voltage-sensitive dye imaging with the sophisticated mathematical methods of principal component analysis and linear discriminant analysis. A candidate key neuron highlighted by these analyses (neuron 208) could selectively bias the decision to swim or crawl.

Bt Receptor Defines Specificity

The Bt toxin, a crystalline protein produced by the soil-borne bacterium *Bacillus thuringiensis*, is used to control insect pests in agriculture. After the toxin is ingested by insect larvae, the toxin damages the gut of susceptible insects. **Griffitts *et al.*** (p. 922) examined the mode of action of Bt. Several genes known to control resistance to the Bt toxin encode enzymes that synthesize a set of glycolipids found in nematodes and insects. These glycolipids function as the receptor for the Bt toxin explaining why the toxic effects of Bt are limited to nematodes and insects.

Natural Brominated Bioaccumulators

Halogenated organic compounds can accumulate in animal tissues, in some cases with potentially toxic consequences. Some of these, such as the polybrominated diphenyl ethers (PBDEs) used as flame retardants, have industrial origins. The origins of some classes of bioaccumulating compounds, such as methoxylated polybrominated diphenyl ethers (MeO-BDEs), have been uncertain. **Teuten *et al.*** (p. 917) extracted more than 10 kilograms of blubber from a fatally stranded True's

CONTINUED ON PAGE 813

Published by AAAS

beaked whale, and isolated MeO-BDEs at 99% purity for radiocarbon analysis, which reliably distinguishes carbon of ancient and recent origin. The carbon content of MeO-BDEs was overwhelmingly recent, indicative of a natural rather than industrial origin for these compounds.

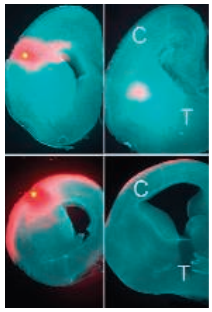
Endangered Ginseng?

Ginseng is a highly valued understory forest plant that is widespread in eastern North America, although at low population density. It has many uses in traditional Asian medicine and strong cultural ties to Appalachian communities. Population viability analyses carried out by **McGraw and Furedi** (p. 920; see the news story by **Stokstad**) suggest that high rates of browsing by burgeoning populations of white-tailed deer threaten to cause extinction of most, if not all, wild American ginseng populations within a century. The white-tailed deer represents a keystone species, with large and cascading effects on the natural community. Loss of the wild populations of ginseng and other potentially valuable understory herbs would have significant economic and cultural consequences.

Earliest Influences

The two main lineages of T cells to emerge from the thymus are distinguished by the T cell receptors that they carry, either $\alpha\beta$ or $\gamma\delta$, which confer distinctive functional properties on each cell type. Within the thymus, the development of the two lineages has been thought to occur independently. **Silva-Santos *et al.*** (p. 925, published online 9 December 2004; see the Perspective by **Rothenberg**) now show that the features peculiar to $\gamma\delta$ T cells are not generated autonomously but are conferred directly on the cells by their immature $\alpha\beta$ thymic counterparts. This process required signaling via a pathway already known to be essential for lymphoid organogenesis and generating effective immune responses. Thus, the developmental interaction between two lineages of T cell imparts fundamental features on one of the cell types.

SADly Promoting Neuronal Polarity



As neurons wire together networks of communication, they need to know not only which other neurons to connect to, but in which direction they should send signals. Such polarity within a single neuron is reflected by its morphology: multiple short dendrites receive signals, and the single longer axon sends signals. **Kishi *et al.*** (p. 929) examined the role of SAD kinases, relatives of nematode synaptic differentiation regulators, in establishing neuronal polarity. Neurons lacking SAD kinases did not polarize to produce morphologically and functionally distinct axons and dendrites.

Overcoming Stress

Diverse human diseases such as viral infections, diabetes, and neurodegeneration are characterized at the cellular level by an inability of the endoplasmic reticulum (ER) to fold proteins properly, resulting in the onset of "ER stress." Uncorrected ER stress activates apoptotic cell death pathways, and it has been hypothesized that these pathways might be manipulated for therapeutic benefit. In a chemical screen, **Boyce *et al.*** (p. 935) identified a small molecule (salubrinal) that protects cells from ER stress–induced apoptosis. Salubrinal selectively inhibited the dephosphorylation of eukaryotic translation initiation factor α (eIF2α), and inhibited herpesvirus replication. Thus, eIF2α may be a valuable drug target for diseases involving ER stress.

Putting the Methyl in Plant MicroRNAs

MicroRNAs (miRNAs), ~22 nucleotide RNAs encoded in the genomes of both plants and animals, have the potential to regulate the expression of a diverse array of genes. Numerous factors modulate miRNA function, for example, *Arabidopsis* mutants of HEN1 show reduced miRNA abundance, as well as miRNA size heterogeneity. **Yu *et al.*** (p. 932) now show that HEN1 methylates miRNAs on the ribose of their last nucleotide. Methylation plays an important role in ribosomal RNA function and stabilizes exogenously introduced small interfering RNAs. It is likely that many, and possibly all, plant miRNAs are similarly methylated, whereas present evidence suggests that animal miRNAs are not methylated.

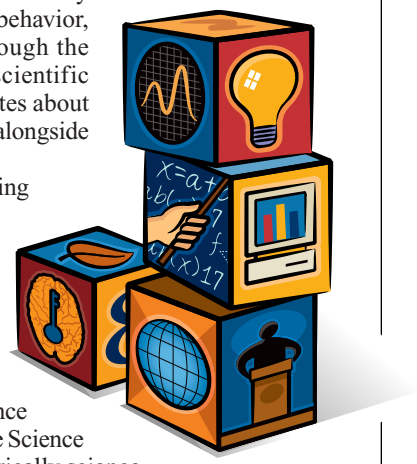
CREDIT: KISHI *ET AL.*

Turn to page 943

Where Science Meets Society

The theme for next week's American Association for the Advancement of Science (AAAS) annual meeting, "The Nexus: Where Science Meets Society," reminds us of many events of the past few years that suggest that the relationship between science and society is undergoing significant stress. Some members of the public are finding certain lines of scientific research and their outcomes disquieting, while others challenge the kind of science taught in schools. This disaffection and shift in attitudes predict a more difficult and intrusive relationship between science and society than we've enjoyed in the recent past.

Examples of these strains in the relationship include sharp public divisions about therapeutic or research cloning and stem cell research. Although many understand the potential benefits of such research, they also are troubled about scientists working so close to what they see as the essence and origins of human life. Last year, ideology came dangerously close to publicly trumping science when the U.S. Congress failed by only two votes to defund a set of grants from the National Institutes of Health on sexual behavior, HIV/AIDS, and drug abuse that made religious conservatives uncomfortable, even though the research was critical to solving major public health problems. And, of course, the scientific community is enmeshed in a continuing battle to keep the nature of science clear in debates about whether schools should be allowed to teach non–science-based "intelligent design theory" alongside evolution in science classrooms.



The common thread linking these examples is that science and its products are intersecting more frequently with certain human beliefs and values. As science encroaches more closely on heavily value-laden issues, members of the public are claiming a stronger role in both the regulation of science and the shaping of the research agenda.

To many, this appears to be a new dimension of the science/society relationship (in truth, it may be a recurrent dimension, because the same issues have been prominent at other historical moments). We've been used to having science and technology evaluated primarily on the basis of potential risks and benefits. However, our recent experience suggests that a third, values-related dimension will influence the conduct and support of science in the future. Taizo Nishimuro, chairman of the board at Toshiba Corporation, suggested at the Science and Technology in Society Forum in Kyoto, Japan, in November 2004 that whereas historically science and technology have changed society, society now is likely to want to change science and technology, or at least to help shape their course.

For many scientists, any such overlay of values on the conduct of science is anathema to our core principles and our historic success. Within the limits of the ethical conduct of science with human or animal subjects, many believe that no scientifically answerable question should be out of bounds. Bringing the power of scientific inquiry to bear on society's most difficult questions is what we have done best, and that often means telling the world things that it might not initially like.

Independence and objectivity in the shaping and conduct of science have been central to our successes and our ability to serve society. Still, our recent experiences suggest that the values dimension is here to stay, certainly for a while, and that we need to learn to work within this new context. Protesting the imposition of value-related constraints on science has been the usual response, but it doesn't work because it doesn't resonate with the public.

An alternative is to adopt a much more inclusive approach that engages other communities assertively in discussing the meaning and usefulness of our work. We should try to find common ground through open, rational discourse. We have had some success with programs such as the National Human Genome Research Institute's Ethical, Legal and Social Implications program. Another example is the AAAS's Dialogue on Science, Ethics, and Religion, which brings scientists together with religious leaders and ethicists to discuss scientific advances and how they relate to other belief and value systems.

Simply protesting the incursion of value considerations into the conduct and use of science confirms the old adage that insanity is doing the same thing over and over and expecting a different outcome. Let's try some diplomacy and discussion and see how that goes for a change.

Alan I. Leshner
Chief Executive Officer, American Association for the Advancement of Science
Executive Publisher, Science

10.1126/science.1110260

EDITORS' CHOICE

edited by Gilbert Chin

ECOLOGY/EVOLUTION

A Forest Sere

Tropical rainforests, despite their locations, can suffer from drought, and during severe droughts, a rainforest can even become susceptible to fire. Evidence of past forest fires, in the form of charcoal deposits, can be found in many parts of the humid tropics, but there has been little documentation of the effects of such catastrophic disturbances on the ecology of tree species.

Van Nieuwstadt and Sheil have examined the effects of drought and fire in a lowland rainforest in East Kalimantan, Indonesia, by censusing live and dead trees in adjacent burned and unburned areas. The drought of 1997–1998, one of the most severe ever in a tropical rainforest, was followed by fire. The consequences of the drought were more pronounced in the larger, mature trees: Nearly half of the trees with trunk diameter >80 cm were lost, whereas less than one-quarter of trees <20 cm in diameter died. In contrast, fire killed smaller saplings disproportionately: Almost no individuals <10 cm in diameter survived. Some species (particularly dipterocarp and palm) withstood fire better than others. In sum, drought and fire both reduce biomass, alter patterns of forest dynamics by removing reproductive individuals and regenerating saplings, and change the relative abundances of species, but do so in different ways. — AMS

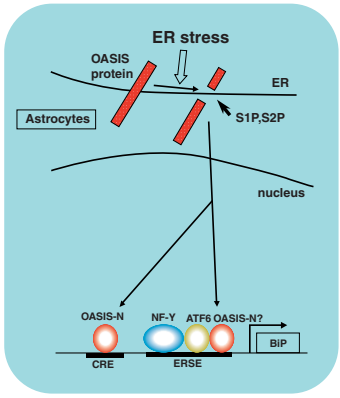
Views from within (inset) and above the forest, showing the effect of drought on larger trees.

J. Ecol. **93**, 191 (2005).

CELL BIOLOGY

Astrocytes and Stress

Eukaryotic cells sense stressful conditions, such as the accumulation of abnormal proteins, in their endoplasmic reticulum (ER) by means of the aptly named unfolded protein response (UPR). As a protective mechanism, the UPR system activates the expression of damage-control proteins, such as the ER protein-folding chaperonin BiP. Kondo *et al.* have determined that astrocytes of the central nervous system employ an ER stress transducer called old astrocyte specifically induced substance (OASIS). OASIS is an ER transmembrane protein in the same transcription factor family as CREB/ATF. When astrocytes were treated with agents that disrupt protein glycosylation or calcium homeostasis in the ER, OASIS was cleaved, and its N-terminal domain moved into the nucleus. This fragment stimulated transcription by activating a promoter with known ER stress-responsive elements (ERSEs). ER stress induced OASIS expression in astrocytes but not in neurons or fibroblasts. Knockdown of OASIS expression reduced the expression of BiP, whereas OASIS overexpression conferred resistance to cell death in response to ER stress. Thus, astrocytes may utilize a cell type–specific mechanism to survive stress induced by ischemic or hypoxic conditions. — LDC



Pathway for OASIS activation of the UPR.

Nature Cell Biol. 10.1038/ncb1213 (2005).

CHEMISTRY

Xe as a Ligand

For more than 20 years, liquid xenon (Xe) has been used as a solvent for studying highly reactive transition metal compounds that attack solvents usually thought of as inert, such as alkanes and fluorocarbons. Nevertheless, infrared spectroscopy showed that in some cases, the Xe reacted transiently with the metal centers, binding with an enthalpy comparable to that of a hydrogen bond. Ball *et al.* have characterized a rhenium (Re)–Xe linkage directly by low-temperature nuclear magnetic resonance (NMR) spectroscopy. They prepared a Xe solution of $(^{i}PrCp)Re(CO)_2PF_3$ (where $^{i}PrCp$ is isopropylcyclopentadienyl) and induced CO loss by ultraviolet irradiation through an optical fiber inserted into the sample probe. The appearance of a ^{129}Xe NMR signal, shifted more than 700 parts per million upfield from the free solvent, confirmed that a Xe atom was coordinated to the unsaturated Re center, and further evidence came from nuclear spin coupling of the bound Xe to the PF_3 ligand, observed via ^{31}P and ^{19}F NMR spectra. The compound persists for hours in liquid Xe at –110°C. — JSY

Proc. Natl. Acad. Sci. U.S.A. **102**, 1853 (2005).

IMMUNOLOGY

Arresting Connections

Our T cell repertoire is individually tailored by positive selection, during which developing thymocytes are vetted for their ability to interact appropriately with self peptides bound to major histocompatibility complex proteins. Using two-photon microscopy, Bhakta *et al.* scrutinized the calcium concentration and motility of thymocytes undergoing positive selection. To maintain the intricate thymic stromal environment, thymocytes were labeled with a dye and introduced into slices of intact thymic tissue. Under conditions in which positive selection did not take place, thymocytes wandered about in much the same way as naïve lymphocytes have been observed to do in lymph nodes. However, this behavior altered in positive selection environments, with thymocytes slowing down considerably and prolonging their interactions with cells of the thymic stroma. Furthermore, these thymocytes displayed increased oscillations of intracellular calcium, indicative of cellular activation. Interruption of Ca^{2+} signaling

CONTINUED ON PAGE 819

was sufficient to restore motility to the thymocytes, suggesting that Ca^{2+} is induced to promote positive selection, most likely by modifying the expression of genes that favor interactions with the thymic stroma. — SJS

Nature Immunol. **6**, 143 (2005).

MATERIAL SCIENCE

Primarily White

For organic light-emitting devices (OLEDs), white light emission has been achieved through the complex and tailored fabrication of multilayer devices either by evaporative or spin coating deposition, or by the blending of two blue-light emitters whose interactions give rise to an exciplex state. In all of these cases, the purity of the white light depends on the quality and concentration of the various species, and generally is a function of the applied voltage.

Mazzeo *et al*. have fabricated an OLED that requires only a single layer of material to generate white light by using an oligothiophene compound. As single molecules in solution, this compound has an intrinsic blue-green emission, whereas in the solid phase, it also produces a red-shifted emission, as crosslinked dimers form. Optical measurements on thiophene compounds that did not form dimers did not show a red-shifted emission spectrum. When wired into a device, the oligothiophene showed electroluminescent emission spectra similar to its photoluminescence, but with a more intense red-shifted peak, leading to the emission of white light (superposed blue-green and red emissions). The intensity of the output in air was similar to that of the best multilayer OLEDs, indicating that this material may find use in general lighting applications. — MSL

Adv. Mater. **17**, 34 (2005).

BIOMEDICINE

Being Sensible About Cholesterol

As a recent advertising campaign reminds us, high cholesterol cannot be blamed solely on our unhealthy diets—the genes we inherit play a role as well. Analyzing a large multiethnic population in Texas, Cohen *et al*. found that individuals with exceptionally low levels of low-density lipoprotein cholesterol (LDL-C), or bad cholesterol, were far more likely than average to carry nonsense mutations in a gene called *PCSK9*; these mutations were found almost exclusively in African-Americans.

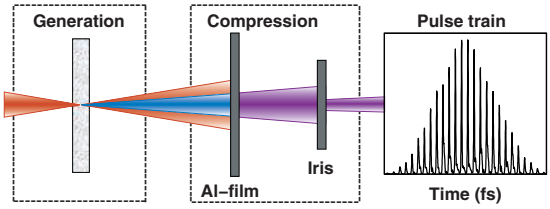
Missense mutations in *PCSK9* had previously been identified as the cause of a rare inherited disorder characterized by extremely high cholesterol levels. The *PCSK9* product is a serine protease (proprotein convertase subtilisin kexin 9), and an independent study of cultured human liver cells describes its role in cholesterol metabolism. By comparing the properties of cells overexpressing the wild type and a catalytically inactive form of the protease, Maxwell *et al*. conclude that PCSK9 accelerates the degradation of a protein that is a key determinant of plasma LDL-C levels, the LDL receptor. — PAK

Nature Genet. **37**, 161 (2005); *Proc. Natl. Acad. Sci. U.S.A.* **102**, 2069 (2005).

PHYSICS

Getting Attosecond Pulses into Shape

The ionization of atoms by intense infrared laser pulses produces light that spans the frequency spectrum from the ultraviolet to soft x-rays. Because this broadband output is made up of many harmonics of the central emission frequency, it should be possible to produce light pulses of several tens of attoseconds in duration. However, not being able to harness the output light has meant that the pulses tend to be several hundreds of attoseconds instead. López-Martens *et al*. show that by compressing and spatially filtering the output light, they can effectively control the phase and amplitude of the attosecond pulses and reduce the length of the pulses to just 170 attoseconds. Such controlled pulses and trains of pulses should provide the precision tools necessary to probe some of the fastest electronic processes, such as the dynamics of atomic excitations and electron orbits. — ISO



Schematic of the experimental design.

Phys. Rev. Lett. **94**, 033001 (2005).

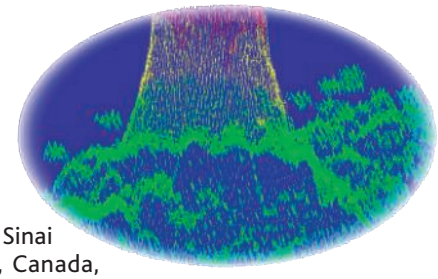
CREDITS: LÓPEZ-MARTENS *ET AL., PHYS. REV. LETT.* **94**, 033001 (2005)

NetWatch

edited by Mitch Leslie

TOOLS

Genomics Workshop



This site from Mount Sinai Hospital in Toronto, Canada, supplies a dozen free tools and programs for sorting and analyzing data about genes and proteins. One popular offering is BIND, which identifies interactions between a specific molecule and others, based on information gleaned from papers, databases, and researcher contributions. Fire up TraDES to predict the three-dimensional structure of a protein from its amino acid sequence (above, a diagram showing possible positions of one amino acid). You can also compare the genomes of multiple species or pinpoint small molecules that bind to particular proteins.

www.blueprint.org

EDUCATION

Small-Screen Science

Cable TV offers more than just braying pundits and endless *Law and Order* reruns. Many cable systems also carry the ResearchChannel, which bills itself as "the C-SPAN of scientific and medical research" and broadcasts educational programming 24 hours a day. You can also catch the channel's mix of lectures, interviews, forums, and lab visits at its Web site. ResearchChannel shows come from more than 25 universities, the National Institutes of Health, and other sources. Recent viewers might have watched a tutorial for researchers on designing clinical trials or a discussion aimed at a general audience about the end of the universe, which featured Nobel laureate Leon Lederman and other experts. You can also call up an archive of some 1700 past programs.

www.researchchannel.org

RESOURCES

Shot Into Space

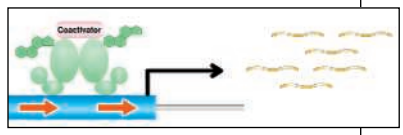
From Japanese red-bellied newts to pepper plants to jellyfish, a bevy of organisms has undergone scientific scrutiny while flying on board NASA spacecraft. The agency's Life Sciences Data Archive stows descriptions of more than 900 of these studies. For example, the newts rode the space shuttle in the summer of 1994 to help researchers determine how weightlessness affects the development of a gravity-sensing structure in the inner ear. To delve deeper, you can download raw data or peruse charts, tables, and other summaries. Here, the astrochimp Ham gets a warm reception after returning from a 16-minute suborbital trip in 1961.

lsda.jsc.nasa.gov

COMMUNITY SITE

Receptor Roundup

Many hormones attach to receptors on the cell surface, but molecules such as testosterone and estrogen link up with so-called nuclear receptors within the cell. The combination then latches onto DNA, turning genes off or on. At the Nuclear Receptor Signaling Atlas (NURSA), molecular biologists, drug designers, and other researchers can uncover information about nuclear receptors, which can go awry in prostate and breast cancer and in conditions such as obesity.



The site's central database describes 49 receptors and, for some, supplies measurements of their messenger RNA levels at different times of the day and in various tissues. The database will grow to include receptor DNA sequences and crystal structures, says site editor Neil McKenna of Baylor College of Medicine in Houston, Texas. NURSA also offers a tutorial on the discovery of nuclear receptors and their interactions with other molecules, such as the coactivators and corepressors that ramp up or hinder their activity. Above, a receptor, hormone, and coactivator amalgam gloms onto a gene. Visitors can also join a discussion forum or browse the site's new, free-access journal.

www.nursa.org

WEB MAGAZINE

Spotlighting Africa's Research

Africa is known as a great place for fieldwork on, say, human origins or cheetah behavior, but you rarely hear about the continent's own researchers. *Science in Africa*, a popular Web magazine edited by a graduate student and a biotechnology lecturer in South Africa, aims to spread the word. The 4-year-old publication features articles about African research in areas such as ecology and genetics, often written by African scientists, and posts news stories on important issues for the continent. One recent feature whisked readers into the bush to check on possible overharvesting of the mopane worm, a tasty caterpillar prized as a source of protein in southern Africa. Another story looked at attempts to predict Africa's version of El Niño.

www.scienceinafrica.co.za/index.htm

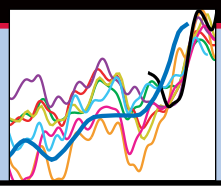
Send site suggestions to netwatch@aaas.org. Archive: www.sciencemag.org/netwatch

CREDITS (TOP TO BOTTOM): THE BLUEPRINT INITIATIVE; NURSA; NASA

Published by AAAS

This Week

PAGE 826
Deer and ginseng



827
Warming and the "hockey stick"

CONFLICT OF INTEREST

NIH Chief Clamps Down on Consulting and Stock Ownership

Under intense pressure from outside, National Institutes of Health (NIH) Director Elias Zerhouni last week issued unexpectedly strict ethics rules intended to "[preserve] the public's trust" in the agency. Congressional critics, who have been troubled by revelations about apparent conflicts of interest among some senior NIH scientists, praised the new rules, but many NIH staff members are outraged, calling them punitive and draconian. Under the rules, which took effect on 3 February, all NIH staff are barred from outside paid or unpaid consulting for drug and medical companies and even nonprofit organizations. All 17,500 employees will also have to sell or limit their stock in biotech and drug companies.

As recently as December, Zerhouni said some industry consulting should be allowed. But pressures from Congress and the difficulty of forging workable rules led him to decide on a "clean break." As he told employees on 2 February, "this issue was standing between the prestigious history of NIH and its future." Leaders in the biomedical research community say the harsh steps were unavoidable as questions continued to arise about a handful of intramural researchers who apparently violated existing ethics rules. "The ground was cut out from under Zerhouni. His hand was forced because of the behaviors of a few," says David Korn, a former medical school dean at Stanford University who is now a senior vice president at the Association of American Medical Colleges (AAMC).

Zerhouni concedes that he had no choice. Ethics concerns first surfaced in 2003 when Congress inquired about cash awards received by then–National Cancer Institute (NCI) director Richard Klausner. NCI ethics officials had approved the awards, but a House subcommittee suggested that gifts from grantee institutions posed a conflict of interest. Then came a December 2003 article in the *Los Angeles Times* reporting that several top scientists had received hundreds of thousands of dollars in payments from industry and raising questions about conflicts of interest. (Former director Harold Varmus had loosened the rules on consulting in 1995 to make them more consistent with those of academia and help recruit talent to the intramural program.)

As Congress began investigating, Zerhouni conferred with an outside panel and proposed new limits. But last summer more problems arose: According to data from drug companies, several dozen employees hadn't told NIH about their consulting activities. "It was like getting shot in the back by your own troops," says Zerhouni.

He then proposed a 1-year moratorium on consulting, but again, more concerns emerged in the press: Some researchers were apparently paid to endorse particular drugs. The final straw came when the Senate appropriations committee suggested that failure to take strong measures could become "a basis for a cut" in NIH's budget, Zerhouni says.

Clean slate. Zerhouni decided only a ban on consulting could resolve past problems.

The new rules, a 96-page Department of Health and Human Services (HHS) interim regulation, ban all compensated and uncompensated consulting or speaking for drug, biotech, and medical-device companies, health care providers, institutions with NIH grants, and even professional societies (see table). NIH scientists can still receive payments for teaching courses, editing and writing for peer-reviewed publications, and practicing medicine. NIH researchers can also continue some activities, such as serving on a society's board, if their supervisor approves it as official duty.

The rules also prohibit senior staff members who file public or confidential financial disclosure reports—about 6000, including many researchers—and their spouses and minor children from owning biotech or drug company stock, a rule followed only by regulatory agencies such as the Food and Drug Administration and the Securities and Exchange Commission. Other NIH employees, such as secretaries and technicians, can keep no more than $15,000 in related stock. The rules restrict cash scientific awards to $200, except for employees below senior level with no business with the donor. (There is an exception for a few awards such as the Nobel Prize.) So far only about 100 awards have been deemed "bona fide."

Although HHS will collect comments on the "interim" regulation for 60 days and evaluate it after a year, the rules will stand unless they are clearly harming recruitment and retention, Zerhouni said. NIH employees will have just 30 to 90 days to end their outside activities and up to 150 days to divest stock.

At an employee meeting last week, staff members reacted angrily to the rules, which one researcher described as "throwing the baby out with the bathwater." The edict to sell off stock, in particular, has hit a nerve: With stock prices low, it could cause "potentially irreparable financial harm," warned one lab chief who, like others, asked a reporter not to use his name. Others questioned the ration- ▶

CREDITS: NIH

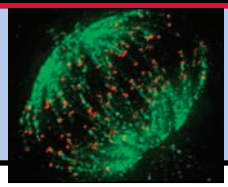
NIH Ethics Rules

Prohibited	Exceptions
No paid or unpaid employment, including consulting, for a drug, biotech, or medical-device company; health care provider; health or science trade, professional, or advocacy group; or research or educational institution with NIH funding.	Can be paid to teach certain university or continuing education courses, write for or edit peer-reviewed publications, or practice medicine part-time. Activities such as society officer may be allowed as official duty.
Financial disclosure filers cannot own financial interests in biotech, drug, or medical-device companies. $15,000 cap for other staff.	Can hold diversified mutual funds, pension or other benefits arising from pre-NIH employment.
Cash awards: Only those on "bona fide" approved list allowed; senior employees and staff with duties involving donor cannot receive cash prize over $200.	Can accept honoraria and travel expenses; full cash prize allowed for major awards such as Nobel. No limits on prizes of little value.

Published by AAAS

832 U.S. budget squeeze

836 SUMO's heavyweight role



840 Taking the measure of Neandertals

ale. NIH Deputy Director Raynard Kington responded that although NIH, unlike FDA, does not regulate companies, its "influence [has become] substantial," citing a drop in the market in December after two large NIH trials using COX-2 inhibitor painkillers were halted for safety reasons.

Scientific groups outside NIH, such as AAMC, generally support the new rules—with caveats. "The nuances and consequences must be watched very, very carefully," says Korn. The Federation of American Societies for Experimental Biology (FASEB) expressed concerns about recruiting as well as possible limits on participating in scientific societies. "It would be a serious loss if those activities were completely curtailed," said FASEB president Paul Kincade.

Thomas Cech, president of the Howard Hughes Medical Institute in Chevy Chase, Maryland, worries that the rules could undermine Zerhouni's goal of translating research into cures. "Medical uses require commercialization. It's not something to be ashamed about. The key thing is to manage to avoid conflict of interest," Cech says.

The new rules seem "like a heavy-handed solution," says Varmus, now president of Memorial Sloan-Kettering Cancer Center in New York City. But thanks to other reforms in the 1990s, "the intramural program is strong, and it can survive," he says. Top scientists will still be attracted to NIH, where they are protected from the vagaries of winning grants in a tight budget climate, he says. "The people who just want to do science will still come here," agrees Robert Nussbaum, a branch chief at the genome institute. But exactly what NIH will look like under some of the most stringent ethics rules in the federal government may not become apparent for several years. **–JOCELYN KAISER**

SCIENTIFIC PUBLISHING

NIH Wants Public Access to Papers 'As Soon As Possible'

Ending months of uncertainty, National Institutes of Health (NIH) Director Elias Zerhouni last week unveiled a policy aimed at making the results of research it funds more freely available. But the announcement has injected a new element of controversy into an already bitter debate. Zerhouni is asking NIH-funded researchers to send copies of manuscripts that have been accepted for publication to a free NIH archive. Researchers will specify when the archive can make them publicly available, but NIH wants that to be "as soon as possible (and within 12 months of the publisher's official date of final publication)." That language has stirred worries that NIH is putting authors on the spot by asking them to challenge publishers' own release dates.

The "public access" policy emerges from a major battle last year. At the request of Congress, NIH in September asked for comment on a proposal to urge its grantees to submit copies of their research manuscripts for posting on NIH's PubMed Central archive 6 months after publication. NIH argued that this would increase public access to research and help it manage research programs. Supporting this plan were librarians, patient advocates, and some scientists who feel that journal prices are too high and that access to research articles should be free. In the other corner, publishers said that free access so soon after publication could bankrupt them and inflict damage on scientific societies dependent on journal income.

After collecting more than 6000 comments from both sides, Zerhouni on 3 February issued a final policy* that states NIH will wait up to 1 year to post the papers, although it "strongly encourages" posting "as soon as possible." This "flexibility" will help protect publishers who believe earlier posting will harm revenues, he says. Norka Ruiz Bravo, NIH deputy director for extramural research, expects that authors "will negotiate" the timing with the publisher rather than relying on the publisher's policy for when articles can be posted. NIH will not track compliance or make public access a condition of accepting an NIH grant, she says: "We have no plans to punish anybody who doesn't follow the policy."

Authors vs. publishers? NIH's Ruiz Bravo urges authors to ask publishers to allow speedy free access to articles.

The policy applies only to original research manuscripts, and authors will send in the final peer-reviewed version accepted for publication. If the author wishes, PubMed Central will incorporate subsequent copy-editing changes to avoid having two slightly different versions of the paper. Alternatively, publishers can have NIH replace the manuscript in PubMed Central with the final published paper.

NIH didn't attempt an economic analysis of the impact on journals, Ruiz Bravo says, because that "would be a major thing." However, the agency argues that because NIH-funded papers make up only 10% of the biomedical research literature, the policy won't put journals out of business; NIH promises to track the impact of the policy through a new advisory group.

Neither side seems satisfied. A group of nonprofit publishers called the D.C. Principles Coalition argues that the $2 million to $4 million per year that NIH estimates it will cost to post 60,000 papers is an unnecessary expense because most nonprofit journals already make papers publicly available in their own searchable archives after a year. "We're concerned about the waste of research dollars," says Martin Frank, executive director of the American Physiological Society in Bethesda, Maryland. Frank also argues that the plan would infringe journals' copyright, and it might not stand up to a legal challenge.

For their part, open-access advocates aren't happy about the "voluntary" aspect or the 12-month timeframe. Whether articles will become available any sooner than they are now "is a big 'if,' " says Sharon Terry, president of the Genetic Alliance and an organizer of the Alliance for Taxpayer Access in Washington, D.C. The request that authors try to have their papers posted as soon as possible puts them "in the untenable position" of trying to please both NIH and their publishers, says the Alliance for Taxpayer Access.

The only group that seems pleased with the wording is the Public Library of Science (PLoS) in San Francisco, California, which ▶

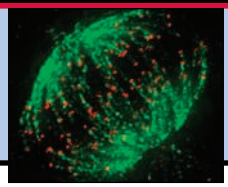
* www.nih.gov/about/publicaccess

CREDIT: NIH

Published by AAAS

832 U.S. budget squeeze

836 SUMO's heavyweight role



840 Taking the measure of Neandertals

ale. NIH Deputy Director Raynard Kington responded that although NIH, unlike FDA, does not regulate companies, its "influence [has become] substantial," citing a drop in the market in December after two large NIH trials using COX-2 inhibitor painkillers were halted for safety reasons.

Scientific groups outside NIH, such as AAMC, generally support the new rules—with caveats. "The nuances and consequences must be watched very, very carefully," says Korn. The Federation of American Societies for Experimental Biology (FASEB) expressed concerns about recruiting as well as possible limits on participating in scientific societies. "It would be a serious loss if those activities were completely curtailed," said FASEB president Paul Kincade.

Thomas Cech, president of the Howard Hughes Medical Institute in Chevy Chase, Maryland, worries that the rules could undermine Zerhouni's goal of translating research into cures. "Medical uses require commercialization. It's not something to be ashamed about. The key thing is to manage to avoid conflict of interest," Cech says.

The new rules seem "like a heavy-handed solution," says Varmus, now president of Memorial Sloan-Kettering Cancer Center in New York City. But thanks to other reforms in the 1990s, "the intramural program is strong, and it can survive," he says. Top scientists will still be attracted to NIH, where they are protected from the vagaries of winning grants in a tight budget climate, he says. "The people who just want to do science will still come here," agrees Robert Nussbaum, a branch chief at the genome institute. But exactly what NIH will look like under some of the most stringent ethics rules in the federal government may not become apparent for several years. **–JOCELYN KAISER**

SCIENTIFIC PUBLISHING

NIH Wants Public Access to Papers 'As Soon As Possible'

Ending months of uncertainty, National Institutes of Health (NIH) Director Elias Zerhouni last week unveiled a policy aimed at making the results of research it funds more freely available. But the announcement has injected a new element of controversy into an already bitter debate. Zerhouni is asking NIH-funded researchers to send copies of manuscripts that have been accepted for publication to a free NIH archive. Researchers will specify when the archive can make them publicly available, but NIH wants that to be "as soon as possible (and within 12 months of the publisher's official date of final publication)." That language has stirred worries that NIH is putting authors on the spot by asking them to challenge publishers' own release dates.

The "public access" policy emerges from a major battle last year. At the request of Congress, NIH in September asked for comment on a proposal to urge its grantees to submit copies of their research manuscripts for posting on NIH's PubMed Central archive 6 months after publication. NIH argued that this would increase public access to research and help it manage research programs. Supporting this plan were librarians, patient advocates, and some scientists who feel that journal prices are too high and that access to research articles should be free. In the other corner, publishers said that free access so soon after publication could bankrupt them and inflict damage on scientific societies dependent on journal income.

After collecting more than 6000 comments from both sides, Zerhouni on 3 February issued a final policy* that states NIH will wait up to 1 year to post the papers, although it "strongly encourages" posting "as soon as possible." This "flexibility" will help protect publishers who believe earlier posting will harm revenues, he says. Norka Ruiz Bravo, NIH deputy director for extramural research, expects that authors "will negotiate" the timing with the publisher rather than relying on the publisher's policy for when articles can be posted. NIH will not track compliance or make public access a condition of accepting an NIH grant, she says: "We have no plans to punish anybody who doesn't follow the policy."

Authors vs. publishers? NIH's Ruiz Bravo urges authors to ask publishers to allow speedy free access to articles.

The policy applies only to original research manuscripts, and authors will send in the final peer-reviewed version accepted for publication. If the author wishes, PubMed Central will incorporate subsequent copy-editing changes to avoid having two slightly different versions of the paper. Alternatively, publishers can have NIH replace the manuscript in PubMed Central with the final published paper.

NIH didn't attempt an economic analysis of the impact on journals, Ruiz Bravo says, because that "would be a major thing." However, the agency argues that because NIH-funded papers make up only 10% of the biomedical research literature, the policy won't put journals out of business; NIH promises to track the impact of the policy through a new advisory group.

Neither side seems satisfied. A group of nonprofit publishers called the D.C. Principles Coalition argues that the $2 million to $4 million per year that NIH estimates it will cost to post 60,000 papers is an unnecessary expense because most nonprofit journals already make papers publicly available in their own searchable archives after a year. "We're concerned about the waste of research dollars," says Martin Frank, executive director of the American Physiological Society in Bethesda, Maryland. Frank also argues that the plan would infringe journals' copyright, and it might not stand up to a legal challenge.

For their part, open-access advocates aren't happy about the "voluntary" aspect or the 12-month timeframe. Whether articles will become available any sooner than they are now "is a big 'if,' " says Sharon Terry, president of the Genetic Alliance and an organizer of the Alliance for Taxpayer Access in Washington, D.C. The request that authors try to have their papers posted as soon as possible puts them "in the untenable position" of trying to please both NIH and their publishers, says the Alliance for Taxpayer Access.

The only group that seems pleased with the wording is the Public Library of Science (PLoS) in San Francisco, California, which ▶

* www.nih.gov/about/publicaccess

CREDIT: NIH

Published by AAAS

charges authors publication costs and then posts papers immediately upon publication. "We have influence here," says PLoS cofounder Harold Varmus, president of Memorial Sloan-Kettering Cancer Center in New York City. "The journal may say 12 months, but the journal also wants [the] paper. Researchers are going to be voting with their feet."

But that assertion assumes researchers will feel strongly enough to raise the issue with publishers. Virologist Craig Cameron of Pennsylvania State University, University Park, says he will likely rely on the publisher's existing policy even if it's 12 months. "With everything I have to think about on a daily basis, it's not something I would spend a lot of time on," he says. Authors will be asked to send their manuscripts to NIH starting 2 May. –JOCELYN KAISER

ECOLOGY

Ginseng Threatened by Bambi's Appetite

With few natural predators left, deer are running rampant across much of eastern North America and Europe. In addition to damaging crops, raising the risk of Lyme disease, and smashing into cars, white-tailed deer are eating their way through forests. "This is a widespread conservation problem," says Lee Frelich of the University of Minnesota, Twin Cities. Indeed, on page 920, a detailed, 5-year forest survey of ginseng reveals that deer, if not checked, will almost certainly drive the economically valuable medicinal plant to extinction in the wild.

The survey was conducted by James McGraw, a plant ecologist at West Virginia University in Morgantown, and his graduate student Mary Ann Furedi. Ginseng is one of the most widely harvested medicinal plants in the United States; in 2003, 34,084 kilograms were exported, mainly to Asia, where wild ginseng root fetches a premium. Although the plant (*Panax quinquefolius*) ranges from Georgia to Quebec, it is slow-growing and scarce everywhere.

Oh deer. Deer are eating their way through too much ginseng (*inset*).

CREDITS: U.S. FISH AND WILDLIFE SERVICE

To determine the population trends of ginseng, McGraw and Furedi began a census in West Virginia forests. For 5 years, they checked seven populations of wild ginseng every 3 weeks during the spring and summer. They quickly noticed that plants were disappearing. In some places, all of the largest, most fertile plants were gone by mid-August. At first they suspected ginseng harvesters, but the valuable roots were left. Cameras confirmed that deer were at work. The nibbled plants are less likely to reproduce, and after repeated grazing, they die. Indeed, during the study, populations declined by 2.7% per year on average.

McGraw and Furedi then ran a ginseng population viability analysis. By plugging in the sizes of plants in various populations, mortality rates, and other factors, they learned that current ginseng populations must contain at least 800 plants in order to have a 95% chance of surviving for 100 years.

That's bad news. A broader survey they conducted of 36 ginseng populations across eight states revealed that the median size was just 93 plants and the largest was only 406 plants. At the current rate of grazing, all of these populations "are fluctuating toward extinction," McGraw concludes. Even the biggest population has only a 57% chance of surviving this century.

"This paper has high significance because it's one of the first demonstrations of the direct impact of deer browsing on understory plants," says Daniel Gagnon of the University of Quebec, Montreal. And deer eat more than ginseng. "We could lose a lot of understory species in the next century if these browsing rates continue," McGraw says. That in turn could affect birds, small mammals, and other wildlife that rely on these plants.

McGraw and Furedi calculate that browsing rates must be cut in half to guarantee a 95% chance of survival for any of the 36 ginseng populations they surveyed. That has direct management implications, says Donald Waller of the University of Wisconsin, Madison. "We should be encouraging the recovery of large predators like wolves. It also suggests we should be increasing the effectiveness of human hunting" by emphasizing the killing of does rather than bucks, he adds. Such deer-control measures are controversial: Reintroduction of predators like wolves faces logistical as well as political hurdles, for example. Meanwhile, the deer keep munching.

–ERIK STOKSTAD

ScienceScope

Biosafety Lab Fallout in Boston

New revelations about how Boston University handled an incident in which dangerous bacteria sickened three workers last year may hinder BU's plans to build a biosafety level 4 (BSL-4) lab in the city's South End neighborhood (*Science*, 28 January, p. 501).

When news of the infections broke last month, the university said that it had not suspected tularemia as the cause until October. But BU officials admitted last week that they had conducted tests on two workers in August that showed the presence of infectious bacteria. Because they were not convinced that the samples contained tularemia, they waited until a third worker fell ill in the fall before they closed the lab, ran further tests, and informed public health officials. Also last week, Peter Rice, the beleaguered head of the lab where the tularemia incident took place and chief of infectious diseases, resigned from his positions at BU. Opponents of the BSL-4 lab, meanwhile, are pushing a bill in the Massachusetts Senate which would ban such facilities from the state.

–ANDREW LAWLER

Turning Bombs Into Semiconductors

ALMATY, KAZAKHSTAN—Plans are afoot to create what may be the world's first "nuclear technopark" at one of the enduring legacies of the Cold War. The government of Kazakhstan is reviewing an $80 million proposal to establish a technology incubator at the Semipalatinsk Test Site—a territory nearly as big as Israel—in northeastern Kazakhstan where the Soviet Union detonated its first atom and hydrogen bombs. Since the closure of the Central Asian facility in 1992, Kazakh authorities have been trying to secure risky materials such as plutonium-laced soil (*Science*, 23 May 2003, p. 1220).

Looking to convert a liability into a sustainable venture, the former test site's physicist-caretakers have drafted plans to build an electron accelerator, a gamma irradiator, and other facilities for producing everything from medical radioisotopes to semiconductors. If the government approves the plan and kicks in the start-up money, the technopark would then use tax exemptions and other incentives to entice commercial partners from Kazakhstan and abroad. A decision is due by the end of the month.

–RICHARD STONE

Published by AAAS

charges authors publication costs and then posts papers immediately upon publication. "We have influence here," says PLoS cofounder Harold Varmus, president of Memorial Sloan-Kettering Cancer Center in New York City. "The journal may say 12 months, but the journal also wants [the] paper. Researchers are going to be voting with their feet."

But that assertion assumes researchers will feel strongly enough to raise the issue with publishers. Virologist Craig Cameron of Pennsylvania State University, University Park, says he will likely rely on the publisher's existing policy even if it's 12 months. "With everything I have to think about on a daily basis, it's not something I would spend a lot of time on," he says. Authors will be asked to send their manuscripts to NIH starting 2 May. –JOCELYN KAISER

ECOLOGY

Ginseng Threatened by Bambi's Appetite

With few natural predators left, deer are running rampant across much of eastern North America and Europe. In addition to damaging crops, raising the risk of Lyme disease, and smashing into cars, white-tailed deer are eating their way through forests. "This is a widespread conservation problem," says Lee Frelich of the University of Minnesota, Twin Cities. Indeed, on page 920, a detailed, 5-year forest survey of ginseng reveals that deer, if not checked, will almost certainly drive the economically valuable medicinal plant to extinction in the wild.

The survey was conducted by James McGraw, a plant ecologist at West Virginia University in Morgantown, and his graduate student Mary Ann Furedi. Ginseng is one of the most widely harvested medicinal plants in the United States; in 2003, 34,084 kilograms were exported, mainly to Asia, where wild ginseng root fetches a premium. Although the plant (*Panax quinquefolius*) ranges from Georgia to Quebec, it is slow-growing and scarce everywhere.

Oh deer. Deer are eating their way through too much ginseng (*inset*).

CREDITS: U.S. FISH AND WILDLIFE SERVICE

To determine the population trends of ginseng, McGraw and Furedi began a census in West Virginia forests. For 5 years, they checked seven populations of wild ginseng every 3 weeks during the spring and summer. They quickly noticed that plants were disappearing. In some places, all of the largest, most fertile plants were gone by mid-August. At first they suspected ginseng harvesters, but the valuable roots were left. Cameras confirmed that deer were at work. The nibbled plants are less likely to reproduce, and after repeated grazing, they die. Indeed, during the study, populations declined by 2.7% per year on average.

McGraw and Furedi then ran a ginseng population viability analysis. By plugging in the sizes of plants in various populations, mortality rates, and other factors, they learned that current ginseng populations must contain at least 800 plants in order to have a 95% chance of surviving for 100 years.

That's bad news. A broader survey they conducted of 36 ginseng populations across eight states revealed that the median size was just 93 plants and the largest was only 406 plants. At the current rate of grazing, all of these populations "are fluctuating toward extinction," McGraw concludes. Even the biggest population has only a 57% chance of surviving this century.

"This paper has high significance because it's one of the first demonstrations of the direct impact of deer browsing on understory plants," says Daniel Gagnon of the University of Quebec, Montreal. And deer eat more than ginseng. "We could lose a lot of understory species in the next century if these browsing rates continue," McGraw says. That in turn could affect birds, small mammals, and other wildlife that rely on these plants.

McGraw and Furedi calculate that browsing rates must be cut in half to guarantee a 95% chance of survival for any of the 36 ginseng populations they surveyed. That has direct management implications, says Donald Waller of the University of Wisconsin, Madison. "We should be encouraging the recovery of large predators like wolves. It also suggests we should be increasing the effectiveness of human hunting" by emphasizing the killing of does rather than bucks, he adds. Such deer-control measures are controversial: Reintroduction of predators like wolves faces logistical as well as political hurdles, for example. Meanwhile, the deer keep munching.

–ERIK STOKSTAD

ScienceScope

Biosafety Lab Fallout in Boston

New revelations about how Boston University handled an incident in which dangerous bacteria sickened three workers last year may hinder BU's plans to build a biosafety level 4 (BSL-4) lab in the city's South End neighborhood (*Science*, 28 January, p. 501).

When news of the infections broke last month, the university said that it had not suspected tularemia as the cause until October. But BU officials admitted last week that they had conducted tests on two workers in August that showed the presence of infectious bacteria. Because they were not convinced that the samples contained tularemia, they waited until a third worker fell ill in the fall before they closed the lab, ran further tests, and informed public health officials. Also last week, Peter Rice, the beleaguered head of the lab where the tularemia incident took place and chief of infectious diseases, resigned from his positions at BU. Opponents of the BSL-4 lab, meanwhile, are pushing a bill in the Massachusetts Senate which would ban such facilities from the state.

–ANDREW LAWLER

Turning Bombs Into Semiconductors

ALMATY, KAZAKHSTAN—Plans are afoot to create what may be the world's first "nuclear technopark" at one of the enduring legacies of the Cold War. The government of Kazakhstan is reviewing an $80 million proposal to establish a technology incubator at the Semipalatinsk Test Site—a territory nearly as big as Israel—in northeastern Kazakhstan where the Soviet Union detonated its first atom and hydrogen bombs. Since the closure of the Central Asian facility in 1992, Kazakh authorities have been trying to secure risky materials such as plutonium-laced soil (*Science*, 23 May 2003, p. 1220).

Looking to convert a liability into a sustainable venture, the former test site's physicist-caretakers have drafted plans to build an electron accelerator, a gamma irradiator, and other facilities for producing everything from medical radioisotopes to semiconductors. If the government approves the plan and kicks in the start-up money, the technopark would then use tax exemptions and other incentives to entice commercial partners from Kazakhstan and abroad. A decision is due by the end of the month.

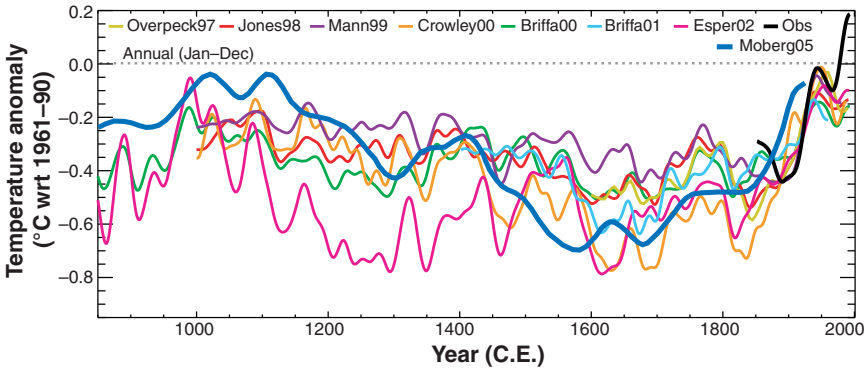
–RICHARD STONE

Published by AAAS

GLOBAL WARMING

Millennium's Hottest Decade Retains Its Title, for Now

The scientific consensus that humans are warming the world stands on three legs, one of which has been getting a workover lately. For a decade, paleoclimatologists have combed through temperature records locked in everything from ancient tree rings to ice cores, yet they've failed to find a natural warming in the past 1000 years as big as that of the past century. That implied that humans and their greenhouse gases were behind the recent warming, as did computer studies of warming patterns and the trend of 20th century warming. But in a soon-to-be-published *Geophysical Research Letters* paper, two researchers attack the recent warming reflected in an iconic paleoclimate record as an artifact of a programming error.



Still no equal. Temperature records recovered from tree rings and other proxies broadly agree that no time in the past millennium has been as warm as recent decades (black).

Even as greenhouse skeptics revel in what they presume is the downfall of one of global warming's most prominent supports, paleoclimatologists have come up with yet another analysis. In a paper published this week in *Nature*, Swedish and Russian researchers present their first entry in the millennial climate sweepstakes. They consider new sorts of measurements and apply a different analytical technique to the data. Their conclusion: Even the surprisingly dynamic climate system doesn't seem to have produced a natural warming as large as that of the past century. "The past couple of decades are still the warmest of the past 1000 years," says climate researcher Philip Jones of the University of East Anglia in Norwich, U.K.

The millennial climate debate has revolved around the "hockey stick" record published in *Nature* by statistical climatologist Michael Mann of the University of Virginia, Charlottesville, and his colleagues in 1998 and revised and extended in 1999. He and his colleagues started with 12 temperature records extracted from, among other things, the width of tree rings, the isotopic composition of ice cores, and the chemical composition of corals—so-called proxies standing in for actual measurements of temperature. They compiled the proxy records and calibrated them against temperatures measured by thermometers in the 20th century. The result was the "hockey stick" curve of Northern Hemisphere temperature over the past millennium. Temperature declined slowly during most of the millennium, creating the long, straight handle of the stick, before rising sharply beginning in the mid–19th century toward the heights of the 1990s, forming the tip of the upturned blade of the stick. Those temperatures handily exceed any temperature of the past millennium.

Two researchers are now saying that the millennial curve doesn't resemble a hockey stick at all. In their latest paper, Stephen McIntyre of Toronto, Canada, a mineral-exploration consultant, and economist Ross McKitrick of the University of Guelph, Canada, make two charges. They claim that "what is almost certainly a computer programming error" in the statistical technique used by Mann and colleagues causes a single record—from ancient bristlecone pine trees of the western United States—to dominate all other records. And the bristlecone pines had a late growth spurt apparently unrelated to rising temperatures, they say. They also charge that Mann's techniques create the appearance of statistical significance in the first half of the millennium where none exists. When McIntyre and McKitrick kicked off a publicity campaign late last month, greenhouse contrarians were gleeful.

Mann calls the McIntyre and McKitrick charges "false and specious." He has been parrying their claims since they responded to his 1998 paper with what he says was an analysis of an inadvertently corrupted data set. The bottom line from the latest go-round, Mann says, is that the same hockey stick appears whether he uses his original technique, variations on it, or a completely dif- ▶

CREDIT: ADAPTED FROM K. R. BRIFFA AND T. J. OSBORN, *SCIENCE* **295** (22 MARCH 2002), AND A. MOBERG *ET AL.*, *NATURE* **322** (10 FEBRUARY 2005)

BIOINFORMATICS

With a Stumble, Microsoft Launches European Research Project

The Microsoft Corp. is about to increase its research presence in Europe. On 2 February, company Chair Bill Gates told a meeting of government leaders in Prague that Microsoft plans to fund several research centers, graduate scholarships, and scientific meetings across Europe, focusing on the interface between computer science and biology, agriculture, and engineering. The venture has been widely welcomed, except for one problem: Its name, the EuroScience Initiative, is already taken.

The initiative's first site will be the Center for Computational and Systems Biology in Trento, Italy. The center will receive up to €15 million over the next 5 years, 60% from national and local governments and 40% from Microsoft. Corrado Priami, a bioinformatics professor at the University of Trento who will head the center, says up to 30 researchers will focus on understanding complex systems such as the chemical communication within a cell and developing tools for biologists and computer designers. Priami says all research results will be made public, and intellectual property will remain with the university, although Microsoft will have an option to exclusively license products that result from the funded research.

Microsoft is reportedly in discussions with universities in Germany, France, and the U.K. and plans to announce several more centers later this year.

As for the name, the EuroScience Association, a group of more than 2000 European scientists founded in 1997, cried foul. The organization, which last year held a European-wide meeting called the EuroScience Open Forum (*Science*, 3 September 2004, p. 1387), also advises the European Union on policy issues, says spokesperson Jens Degett. "If suddenly there is no difference between EuroScience and Microsoft, it will be very damaging" to the group's credibility as an independent organization. In response, Microsoft said it would work with the group to eliminate any misunderstanding and is planning to rename the program. **–GRETCHEN VOGEL**

ferent methodology. Observers have been slow to wade into such turbid statistical waters, citing instead the other half-dozen paleoclimate studies employing a variety of data analyzed using two different types of methodologies. McIntyre, however, sees far too much overlap among analysts and data sets and perceives far too many problems in analyses to be impressed.

Now comes a joint Swedish-Russian effort that clearly breaks away from the pack. Climate researcher Anders Moberg of the University of Stockholm, Sweden, and his colleagues have not participated in previous millennia analyses. Tree rings don't preserve century-scale temperature variations very well, so they added 11 proxy records ranging from cave stalagmites in China to an ice core in northern Canada. They also used a wavelet transform technique for processing the data, a new approach in millennial studies.

Moberg and his colleagues found that temperatures around the hemisphere fell farther during the Little Ice Age of the 17th century than in Mann's reconstruction and rose higher in medieval times. The medieval warmth equaled that of most of the 20th century, but it still did not equal the warmth of 1990 and later.

Moberg's result is only the latest to suggest that the handle of "the hockey stick is not flat," says paleoclimatologist Thomas Crowley of Duke University in Durham, North Carolina. "It's more like a boomerang," he notes. The near end still sticks up—albeit less dramatically—above all else of the past 1000 years.

–RICHARD A. KERR

TOXIC AIR POLLUTANTS

Inspector General Blasts EPA Mercury Analysis

Power plants buying and selling the right to spew toxic mercury from their smokestacks—the mere prospect raises the hackles of environmentalists. But when the U.S. Environmental Protection Agency (EPA) proposed such a cap-and-trade system last year, it argued that it was the most effective way to cut back the 48 tons of mercury, a known neurotoxin, emitted nationwide each year. Last week, the agency came under fire anew—this time from its own Inspector General (IG), who accused EPA officials of deliberately skewing their analyses to burnish the cap-and-trade approach. EPA denies the charges, but environmentalists say the report* will give them a leg up in court if they sue over the final rule.

Up in smoke. Coal-fired power plants account for most mercury emissions in the United States.

Coal-fired power plants are responsible for about 40% of all mercury emissions in the United States, making them the largest single source. Perhaps as much as half spreads considerable distances, while the rest is deposited locally, creating so-called hot spots. The primary route of human exposure is fish consumption, because mercury bioaccumulates in water. Nearly every state has fish consumption advisories, especially for pregnant women, as fetuses are considered most vulnerable.

No federal rules on mercury from power plants are in place yet, although EPA determined in 2000 that regulation was "appropriate and necessary." Under existing law, there is only one way to regulate a hazardous air pollutant like mercury (as opposed to less dangerous pollutants). This so-called MACT (maximum achievable control technology) approach requires all polluters to meet an air standard based on the average emissions of the cleanest 12% of power plants.

While calculating the MACT, EPA became enamored of pollution-trading approaches, allowed by law for so-called criteria or conventional air pollutants. For instance, the "Clear Skies" legislation, introduced in Congress in June 2002, included a pollution-trading scheme to reduce emissions of sulfur dioxide (SO_2) and nitrogen oxides (NO_x). That's relevant to the mercury debate because the same scrubber technology that can clean up these pollutants can also reduce mercury in some situations, yielding what's called a "cobenefit."

After that bill stalled, EPA proposed a rule in January 2004 that would regulate mercury under a similar cap-and-trade system. The agency claimed that this trading approach would cut emissions by 70% to 15 tons by 2018—apparently a much better bottom line than the MACT approach, which EPA said would lower annual emissions to ▶

* Additional Analyses of Mercury Emissions Needed Before EPA Finalizes Rules for Coal-Fired Electric Utilities. www.epa.gov/oigearth/reports/2005/20050203-2005-P-00003-Gcopy.pdf

CREDIT: PHOTODISC RED/GETTY IMAGES

ScienceScope

EPA to Consider Human Pesticide Tests

The Environmental Protection Agency (EPA) will once again accept data from controversial studies that deliberately dose human volunteers with pesticides.

EPA stopped considering such data in December 2001, after the advocacy organization Environmental Working Group (EWG) challenged them as unethical. A review by the National Academy of Sciences (NAS) recommended that EPA accept the results of certain human tests if they met strict scientific and ethical criteria (*Science*, 27 February 2004, p. 1272). Meanwhile, CropLife America, a Washington, D.C.–based industry trade group, had sued EPA arguing that the moratorium was illegal, and in 2003 a judge agreed.

Now EPA has announced in an 8 February *Federal Register* notice that unless the studies are "fundamentally unethical," it will consider them case by case until new guidelines, including an ethics review board, are in place. That's consistent with the NAS recommendations. Still, EWG's Richard Wiles is upset. "This is the worst possible outcome," he says. "There are no rules, as far as I can tell."

–JOCELYN KAISER

Harvard Creates New Task Forces on Women in Science

A month after making controversial remarks about why men outnumber women in most scientific disciplines (*Science*, 28 January, p. 492), Harvard University president Lawrence Summers last week set up two task forces on campus to change the situation. The first, led by historian Evelyn Hammonds, will work to improve faculty searches and create a senior administrative position for improving gender diversity. The second group, chaired by computer scientist Barbara Grosz, will probe why women are underrepresented.

–YUDHIJIT BHATTACHARJEE

Nascent Reform Bill Criticized

PARIS—French scientists took to the streets last week to protest a government bill designed to boost research by reforming it (*Science*, 7 January, p. 27). The bill hasn't been made public yet, but after reviewing a leaked draft, leading scientists have concluded that it focuses too heavily on applied research. The government has scheduled more meetings with unions and leaders this month, so the bill won't be presented to Parliament until March at the earliest.

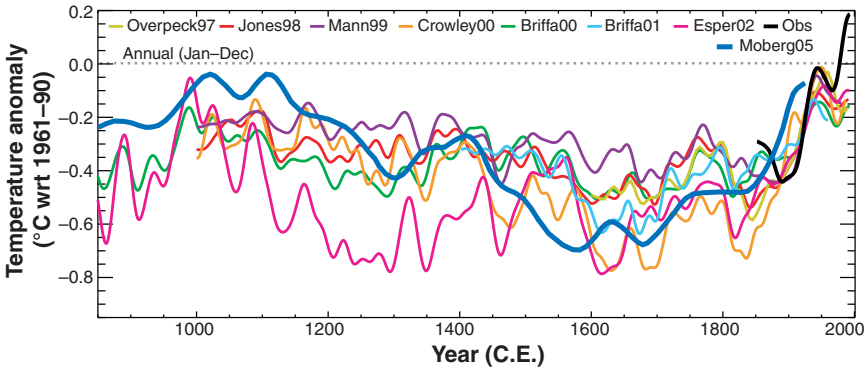
–BARBARA CASASSUS

Published by AAAS

GLOBAL WARMING

Millennium's Hottest Decade Retains Its Title, for Now

The scientific consensus that humans are warming the world stands on three legs, one of which has been getting a workover lately. For a decade, paleoclimatologists have combed through temperature records locked in everything from ancient tree rings to ice cores, yet they've failed to find a natural warming in the past 1000 years as big as that of the past century. That implied that humans and their greenhouse gases were behind the recent warming, as did computer studies of warming patterns and the trend of 20th century warming. But in a soon-to-be-published *Geophysical Research Letters* paper, two researchers attack the recent warming reflected in an iconic paleoclimate record as an artifact of a programming error.



Still no equal. Temperature records recovered from tree rings and other proxies broadly agree that no time in the past millennium has been as warm as recent decades (black).

Even as greenhouse skeptics revel in what they presume is the downfall of one of global warming's most prominent supports, paleoclimatologists have come up with yet another analysis. In a paper published this week in *Nature*, Swedish and Russian researchers present their first entry in the millennial climate sweepstakes. They consider new sorts of measurements and apply a different analytical technique to the data. Their conclusion: Even the surprisingly dynamic climate system doesn't seem to have produced a natural warming as large as that of the past century. "The past couple of decades are still the warmest of the past 1000 years," says climate researcher Philip Jones of the University of East Anglia in Norwich, U.K.

The millennial climate debate has revolved around the "hockey stick" record published in *Nature* by statistical climatologist Michael Mann of the University of Virginia, Charlottesville, and his colleagues in 1998 and revised and extended in 1999. He and his colleagues started with 12 temperature records extracted from, among other things, the width of tree rings, the isotopic composition of ice cores, and the chemical composition of corals—so-called proxies standing in for actual measurements of temperature. They compiled the proxy records and calibrated them against temperatures measured by thermometers in the 20th century. The result was the "hockey stick" curve of Northern Hemisphere temperature over the past millennium. Temperature declined slowly during most of the millennium, creating the long, straight handle of the stick, before rising sharply beginning in the mid–19th century toward the heights of the 1990s, forming the tip of the upturned blade of the stick. Those temperatures handily exceed any temperature of the past millennium.

Two researchers are now saying that the millennial curve doesn't resemble a hockey stick at all. In their latest paper, Stephen McIntyre of Toronto, Canada, a mineral-exploration consultant, and economist Ross McKitrick of the University of Guelph, Canada, make two charges. They claim that "what is almost certainly a computer programming error" in the statistical technique used by Mann and colleagues causes a single record—from ancient bristlecone pine trees of the western United States—to dominate all other records. And the bristlecone pines had a late growth spurt apparently unrelated to rising temperatures, they say. They also charge that Mann's techniques create the appearance of statistical significance in the first half of the millennium where none exists. When McIntyre and McKitrick kicked off a publicity campaign late last month, greenhouse contrarians were gleeful.

Mann calls the McIntyre and McKitrick charges "false and specious." He has been parrying their claims since they responded to his 1998 paper with what he says was an analysis of an inadvertently corrupted data set. The bottom line from the latest go-round, Mann says, is that the same hockey stick appears whether he uses his original technique, variations on it, or a completely dif- ▶

CREDIT: ADAPTED FROM K. R. BRIFFA AND T. J. OSBORN, *SCIENCE* **295** (22 MARCH 2002), AND A. MOBERG *ET AL.*, *NATURE* **322** (10 FEBRUARY 2005)

BIOINFORMATICS

With a Stumble, Microsoft Launches European Research Project

The Microsoft Corp. is about to increase its research presence in Europe. On 2 February, company Chair Bill Gates told a meeting of government leaders in Prague that Microsoft plans to fund several research centers, graduate scholarships, and scientific meetings across Europe, focusing on the interface between computer science and biology, agriculture, and engineering. The venture has been widely welcomed, except for one problem: Its name, the EuroScience Initiative, is already taken.

The initiative's first site will be the Center for Computational and Systems Biology in Trento, Italy. The center will receive up to €15 million over the next 5 years, 60% from national and local governments and 40% from Microsoft. Corrado Priami, a bioinformatics professor at the University of Trento who will head the center, says up to 30 researchers will focus on understanding complex systems such as the chemical communication within a cell and developing tools for biologists and computer designers. Priami says all research results will be made public, and intellectual property will remain with the university, although Microsoft will have an option to exclusively license products that result from the funded research.

Microsoft is reportedly in discussions with universities in Germany, France, and the U.K. and plans to announce several more centers later this year.

As for the name, the EuroScience Association, a group of more than 2000 European scientists founded in 1997, cried foul. The organization, which last year held a European-wide meeting called the EuroScience Open Forum (*Science*, 3 September 2004, p. 1387), also advises the European Union on policy issues, says spokesperson Jens Degett. "If suddenly there is no difference between EuroScience and Microsoft, it will be very damaging" to the group's credibility as an independent organization. In response, Microsoft said it would work with the group to eliminate any misunderstanding and is planning to rename the program. **–GRETCHEN VOGEL**

ferent methodology. Observers have been slow to wade into such turbid statistical waters, citing instead the other half-dozen paleoclimate studies employing a variety of data analyzed using two different types of methodologies. McIntyre, however, sees far too much overlap among analysts and data sets and perceives far too many problems in analyses to be impressed.

Now comes a joint Swedish-Russian effort that clearly breaks away from the pack. Climate researcher Anders Moberg of the University of Stockholm, Sweden, and his colleagues have not participated in previous millennia analyses. Tree rings don't preserve century-scale temperature variations very well, so they added 11 proxy records ranging from cave stalagmites in China to an ice core in northern Canada. They also used a wavelet transform technique for processing the data, a new approach in millennial studies.

Moberg and his colleagues found that temperatures around the hemisphere fell farther during the Little Ice Age of the 17th century than in Mann's reconstruction and rose higher in medieval times. The medieval warmth equaled that of most of the 20th century, but it still did not equal the warmth of 1990 and later.

Moberg's result is only the latest to suggest that the handle of "the hockey stick is not flat," says paleoclimatologist Thomas Crowley of Duke University in Durham, North Carolina. "It's more like a boomerang," he notes. The near end still sticks up—albeit less dramatically—above all else of the past 1000 years.

–RICHARD A. KERR

TOXIC AIR POLLUTANTS

Inspector General Blasts EPA Mercury Analysis

Power plants buying and selling the right to spew toxic mercury from their smokestacks—the mere prospect raises the hackles of environmentalists. But when the U.S. Environmental Protection Agency (EPA) proposed such a cap-and-trade system last year, it argued that it was the most effective way to cut back the 48 tons of mercury, a known neurotoxin, emitted nationwide each year. Last week, the agency came under fire anew—this time from its own Inspector General (IG), who accused EPA officials of deliberately skewing their analyses to burnish the cap-and-trade approach. EPA denies the charges, but environmentalists say the report* will give them a leg up in court if they sue over the final rule.

Up in smoke. Coal-fired power plants account for most mercury emissions in the United States.

Coal-fired power plants are responsible for about 40% of all mercury emissions in the United States, making them the largest single source. Perhaps as much as half spreads considerable distances, while the rest is deposited locally, creating so-called hot spots. The primary route of human exposure is fish consumption, because mercury bioaccumulates in water. Nearly every state has fish consumption advisories, especially for pregnant women, as fetuses are considered most vulnerable.

No federal rules on mercury from power plants are in place yet, although EPA determined in 2000 that regulation was "appropriate and necessary." Under existing law, there is only one way to regulate a hazardous air pollutant like mercury (as opposed to less dangerous pollutants). This so-called MACT (maximum achievable control technology) approach requires all polluters to meet an air standard based on the average emissions of the cleanest 12% of power plants.

While calculating the MACT, EPA became enamored of pollution-trading approaches, allowed by law for so-called criteria or conventional air pollutants. For instance, the "Clear Skies" legislation, introduced in Congress in June 2002, included a pollution-trading scheme to reduce emissions of sulfur dioxide (SO_2) and nitrogen oxides (NO_x). That's relevant to the mercury debate because the same scrubber technology that can clean up these pollutants can also reduce mercury in some situations, yielding what's called a "cobenefit."

After that bill stalled, EPA proposed a rule in January 2004 that would regulate mercury under a similar cap-and-trade system. The agency claimed that this trading approach would cut emissions by 70% to 15 tons by 2018—apparently a much better bottom line than the MACT approach, which EPA said would lower annual emissions to ▶

* Additional Analyses of Mercury Emissions Needed Before EPA Finalizes Rules for Coal-Fired Electric Utilities. www.epa.gov/oigearth/reports/2005/20050203-2005-P-00003-Gcopy.pdf

CREDIT: PHOTODISC RED/GETTY IMAGES

ScienceScope

EPA to Consider Human Pesticide Tests

The Environmental Protection Agency (EPA) will once again accept data from controversial studies that deliberately dose human volunteers with pesticides.

EPA stopped considering such data in December 2001, after the advocacy organization Environmental Working Group (EWG) challenged them as unethical. A review by the National Academy of Sciences (NAS) recommended that EPA accept the results of certain human tests if they met strict scientific and ethical criteria (*Science*, 27 February 2004, p. 1272). Meanwhile, CropLife America, a Washington, D.C.–based industry trade group, had sued EPA arguing that the moratorium was illegal, and in 2003 a judge agreed.

Now EPA has announced in an 8 February *Federal Register* notice that unless the studies are "fundamentally unethical," it will consider them case by case until new guidelines, including an ethics review board, are in place. That's consistent with the NAS recommendations. Still, EWG's Richard Wiles is upset. "This is the worst possible outcome," he says. "There are no rules, as far as I can tell."

–JOCELYN KAISER

Harvard Creates New Task Forces on Women in Science

A month after making controversial remarks about why men outnumber women in most scientific disciplines (*Science*, 28 January, p. 492), Harvard University president Lawrence Summers last week set up two task forces on campus to change the situation. The first, led by historian Evelyn Hammonds, will work to improve faculty searches and create a senior administrative position for improving gender diversity. The second group, chaired by computer scientist Barbara Grosz, will probe why women are underrepresented.

–YUDHIJIT BHATTACHARJEE

Nascent Reform Bill Criticized

PARIS—French scientists took to the streets last week to protest a government bill designed to boost research by reforming it (*Science*, 7 January, p. 27). The bill hasn't been made public yet, but after reviewing a leaked draft, leading scientists have concluded that it focuses too heavily on applied research. The government has scheduled more meetings with unions and leaders this month, so the bill won't be presented to Parliament until March at the earliest.

–BARBARA CASASSUS

Published by AAAS

only 34 tons by 2008. Industry likes this approach, because it gives power plants more flexibility in the technology they can employ and provides time to cope by slowly tightening the regulations.

Environmentalists and state regulatory agencies were highly critical, charging that the trading system would allow the dirtiest power plants to buy the rights to continue polluting, and mercury would continue to accumulate in toxic hot spots. In April of last year, seven senators asked the IG to investigate.

Now the IG has weighed in, charging in a 3 February report that EPA analyses were intentionally "biased" to make the MACT standard look less effective. Citing internal e-mails, the IG maintains that high-ranking officials had their fingers on the scale during this process: "EPA staff were instructed to develop a MACT standard that would result in national emissions of 34 tons per year" by 2008, the report found. Agency documents show that EPA took several stabs at running the model that produces the MACT standards, first yielding 29 tons, then 27, then finally 31. EPA then adjusted the results of the final run to hit the target.

Why 34 tons? The IG notes that's the same reduction that would be achieved as a cobenefit by simply reducing SO_2 and NO_x under the cap-and-trade rule proposed earlier. Martha Keating of the Clean Air Task Force in Boston, Massachusetts, sees it as an attempt to save industry from any extra costs. She says, and state regulators agree, that power plants could achieve greater reductions under MACT if they were required to install new technology, called activated carbon injection. EPA says it didn't generally consider the effects of this technology for its MACT standard, arguing that it won't be commercially ready by 2008.

The IG recommends that EPA rerun its analyses of the MACT standard and tighten its cap-and-trade proposal, but it can't force the agency to do so. EPA says that its final rule, expected 15 March, will include further details, analyses, and cost-benefit information. Spokesperson Cynthia Bergman maintains that the agency properly created the MACT standard and that the cap-and-trade rule is the better way to go. Meanwhile, Senator Jim Jeffords (D–VT), one of those who signed the request to the IG, called for "extensive oversight hearings into this important health issue and into the process by which this rule was crafted." **–ERIK STOKSTAD**

ASTRONOMY

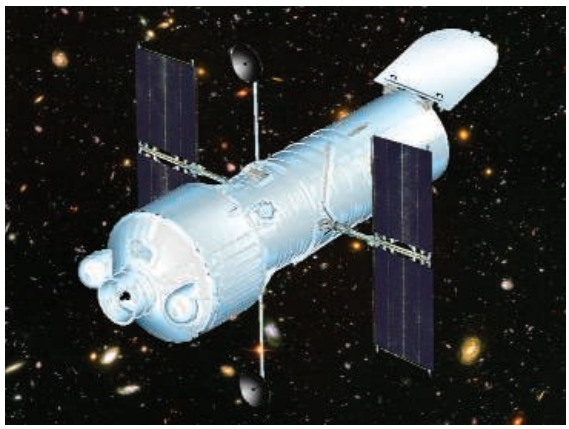
Hearing Highlights Dispute Over Hubble's Future

Scientists, engineers, and politicians are increasingly at odds over what to do with the Hubble Space Telescope. That much was clear at a contentious hearing last week before the House Science Committee, where participants disagreed over whether and how to service the aging spacecraft, what each option would cost, and how to pay for it.

Sean O'Keefe, set to give up his job as NASA administrator, caused a stir last year when he canceled a mission to have astronauts upgrade Hubble's instruments and keep it running until the end of the decade, when the James Webb Space Telescope is slated for launch. After pressure from lawmakers, he suggested that a robotic mission would be a safer bet than sending humans. That proposal, however, was shot down in December by a panel of the National Academy of Sciences, which called the robotic option too complex and costly and urged O'Keefe to reconsider sending astronauts to do the job. The panel also noted that the telescope could fail by 2007, before the robot likely would be ready. This week President George W. Bush requested no funding for a servicing mission in NASA's 2006 budget, a step that seems certain to keep the debate raging.

Representative Sherwood Boehlert (R–NY), who chairs the science committee, called himself an "agnostic" and pleaded with witnesses to "clarify what's at stake." What emerged were the deep divisions among scientists—including those at the same institution.

Astronomer Colin Norman of the Space Telescope Science Institute in Baltimore, Maryland, said the best option is to forgo fixing Hubble in favor of a $1 billion telescope, dubbed Hubble Origins Probe (HOP), that could examine dark energy, dark matter, and planets around other stars in addition to extending Hubble's mission. He noted that Japan has offered to help pay for HOP, which would be launched in 2010. "We must continue with the Hubble adventure," Norman added. The institute's director, Steven Beckwith, also favors completing Hubble's mission. But he wants to do it "as soon as possible," having it fixed by experienced astronauts aboard the shuttle rather than building and launching a new telescope. Other researchers expressed fear that any fix would come at the expense of other science projects.



Follow on. Instead of fixing Hubble, some astronomers are advocating a new telescope, the Hubble Origins Probe.

Joseph Taylor, a Princeton University astronomer who co-chaired the academy's 2000 astronomy study that set long-range priorities, says he opposes any servicing "if it requires major delays or reordering" of future missions. Neither a new telescope nor a servicing mission "should be a higher priority" than the Webb and Constellation-X, another planned NASA telescope, he stated.

Although astronomers are loath to lose Hubble, they also want to protect projects in the decadal study. "We have been playing fast and loose with the process by ignoring our prioritizations," says Alan Dressler of the Carnegie Institution of Washington in Pasadena, California, who did not testify at the hearing. Dressler wants the academy to find out which missions astronomers would be willing to sacrifice to save Hubble.

Louis Lanzerotti, who led the academy's Hubble study, agrees that if science must pay the servicing tab, priorities must be assessed. "If $1 billion is going to come out of some other aspect of NASA's science program, then I would have serious questions" about another Hubble mission—be it a new telescope, a shuttle service, or a robotic effort. But both he and Taylor would back a servicing mission if the money came from elsewhere. Lanzerotti added that NASA's $1 billion estimate doesn't square with the $300 million to $400 million price tag of past shuttle missions: "There is some accounting here which doesn't compute."

For many scientists, NASA's robotic mission is the least attractive option. Lanzerotti, for one, said it would be using an important scientific instrument as "target practice" for new technologies. But Representative Dana Rohrabacher (R–CA) argued that NASA should "push the envelope" by taking the opportunity to develop technologies that could benefit from Bush's plans for space exploration. **–ANDREW LAWLER**

With reporting by Robert Irion.

Published by AAAS

charges authors publication costs and then posts papers immediately upon publication. "We have influence here," says PLoS cofounder Harold Varmus, president of Memorial Sloan-Kettering Cancer Center in New York City. "The journal may say 12 months, but the journal also wants [the] paper. Researchers are going to be voting with their feet."

But that assertion assumes researchers will feel strongly enough to raise the issue with publishers. Virologist Craig Cameron of Pennsylvania State University, University Park, says he will likely rely on the publisher's existing policy even if it's 12 months. "With everything I have to think about on a daily basis, it's not something I would spend a lot of time on," he says. Authors will be asked to send their manuscripts to NIH starting 2 May. –JOCELYN KAISER

ECOLOGY

Ginseng Threatened by Bambi's Appetite

With few natural predators left, deer are running rampant across much of eastern North America and Europe. In addition to damaging crops, raising the risk of Lyme disease, and smashing into cars, white-tailed deer are eating their way through forests. "This is a widespread conservation problem," says Lee Frelich of the University of Minnesota, Twin Cities. Indeed, on page 920, a detailed, 5-year forest survey of ginseng reveals that deer, if not checked, will almost certainly drive the economically valuable medicinal plant to extinction in the wild.

The survey was conducted by James McGraw, a plant ecologist at West Virginia University in Morgantown, and his graduate student Mary Ann Furedi. Ginseng is one of the most widely harvested medicinal plants in the United States; in 2003, 34,084 kilograms were exported, mainly to Asia, where wild ginseng root fetches a premium. Although the plant (*Panax quinquefolius*) ranges from Georgia to Quebec, it is slow-growing and scarce everywhere.

Oh deer. Deer are eating their way through too much ginseng (*inset*).

CREDITS: U.S. FISH AND WILDLIFE SERVICE

To determine the population trends of ginseng, McGraw and Furedi began a census in West Virginia forests. For 5 years, they checked seven populations of wild ginseng every 3 weeks during the spring and summer. They quickly noticed that plants were disappearing. In some places, all of the largest, most fertile plants were gone by mid-August. At first they suspected ginseng harvesters, but the valuable roots were left. Cameras confirmed that deer were at work. The nibbled plants are less likely to reproduce, and after repeated grazing, they die. Indeed, during the study, populations declined by 2.7% per year on average.

McGraw and Furedi then ran a ginseng population viability analysis. By plugging in the sizes of plants in various populations, mortality rates, and other factors, they learned that current ginseng populations must contain at least 800 plants in order to have a 95% chance of surviving for 100 years.

That's bad news. A broader survey they conducted of 36 ginseng populations across eight states revealed that the median size was just 93 plants and the largest was only 406 plants. At the current rate of grazing, all of these populations "are fluctuating toward extinction," McGraw concludes. Even the biggest population has only a 57% chance of surviving this century.

"This paper has high significance because it's one of the first demonstrations of the direct impact of deer browsing on understory plants," says Daniel Gagnon of the University of Quebec, Montreal. And deer eat more than ginseng. "We could lose a lot of understory species in the next century if these browsing rates continue," McGraw says. That in turn could affect birds, small mammals, and other wildlife that rely on these plants.

McGraw and Furedi calculate that browsing rates must be cut in half to guarantee a 95% chance of survival for any of the 36 ginseng populations they surveyed. That has direct management implications, says Donald Waller of the University of Wisconsin, Madison. "We should be encouraging the recovery of large predators like wolves. It also suggests we should be increasing the effectiveness of human hunting" by emphasizing the killing of does rather than bucks, he adds. Such deer-control measures are controversial: Reintroduction of predators like wolves faces logistical as well as political hurdles, for example. Meanwhile, the deer keep munching.

–ERIK STOKSTAD

ScienceScope

Biosafety Lab Fallout in Boston

New revelations about how Boston University handled an incident in which dangerous bacteria sickened three workers last year may hinder BU's plans to build a biosafety level 4 (BSL-4) lab in the city's South End neighborhood (*Science*, 28 January, p. 501).

When news of the infections broke last month, the university said that it had not suspected tularemia as the cause until October. But BU officials admitted last week that they had conducted tests on two workers in August that showed the presence of infectious bacteria. Because they were not convinced that the samples contained tularemia, they waited until a third worker fell ill in the fall before they closed the lab, ran further tests, and informed public health officials. Also last week, Peter Rice, the beleaguered head of the lab where the tularemia incident took place and chief of infectious diseases, resigned from his positions at BU. Opponents of the BSL-4 lab, meanwhile, are pushing a bill in the Massachusetts Senate which would ban such facilities from the state.

–ANDREW LAWLER

Turning Bombs Into Semiconductors

ALMATY, KAZAKHSTAN—Plans are afoot to create what may be the world's first "nuclear technopark" at one of the enduring legacies of the Cold War. The government of Kazakhstan is reviewing an $80 million proposal to establish a technology incubator at the Semipalatinsk Test Site—a territory nearly as big as Israel—in northeastern Kazakhstan where the Soviet Union detonated its first atom and hydrogen bombs. Since the closure of the Central Asian facility in 1992, Kazakh authorities have been trying to secure risky materials such as plutonium-laced soil (*Science*, 23 May 2003, p. 1220).

Looking to convert a liability into a sustainable venture, the former test site's physicist-caretakers have drafted plans to build an electron accelerator, a gamma irradiator, and other facilities for producing everything from medical radioisotopes to semiconductors. If the government approves the plan and kicks in the start-up money, the technopark would then use tax exemptions and other incentives to entice commercial partners from Kazakhstan and abroad. A decision is due by the end of the month.

–RICHARD STONE

Published by AAAS

ferent methodology. Observers have been slow to wade into such turbid statistical waters, citing instead the other half-dozen paleoclimate studies employing a variety of data analyzed using two different types of methodologies. McIntyre, however, sees far too much overlap among analysts and data sets and perceives far too many problems in analyses to be impressed.

Now comes a joint Swedish-Russian effort that clearly breaks away from the pack. Climate researcher Anders Moberg of the University of Stockholm, Sweden, and his colleagues have not participated in previous millennia analyses. Tree rings don't preserve century-scale temperature variations very well, so they added 11 proxy records ranging from cave stalagmites in China to an ice core in northern Canada. They also used a wavelet transform technique for processing the data, a new approach in millennial studies.

Moberg and his colleagues found that temperatures around the hemisphere fell farther during the Little Ice Age of the 17th century than in Mann's reconstruction and rose higher in medieval times. The medieval warmth equaled that of most of the 20th century, but it still did not equal the warmth of 1990 and later.

Moberg's result is only the latest to suggest that the handle of "the hockey stick is not flat," says paleoclimatologist Thomas Crowley of Duke University in Durham, North Carolina. "It's more like a boomerang," he notes. The near end still sticks up—albeit less dramatically—above all else of the past 1000 years.

–RICHARD A. KERR

TOXIC AIR POLLUTANTS

Inspector General Blasts EPA Mercury Analysis

Power plants buying and selling the right to spew toxic mercury from their smokestacks—the mere prospect raises the hackles of environmentalists. But when the U.S. Environmental Protection Agency (EPA) proposed such a cap-and-trade system last year, it argued that it was the most effective way to cut back the 48 tons of mercury, a known neurotoxin, emitted nationwide each year. Last week, the agency came under fire anew—this time from its own Inspector General (IG), who accused EPA officials of deliberately skewing their analyses to burnish the cap-and-trade approach. EPA denies the charges, but environmentalists say the report* will give them a leg up in court if they sue over the final rule.

Up in smoke. Coal-fired power plants account for most mercury emissions in the United States.

Coal-fired power plants are responsible for about 40% of all mercury emissions in the United States, making them the largest single source. Perhaps as much as half spreads considerable distances, while the rest is deposited locally, creating so-called hot spots. The primary route of human exposure is fish consumption, because mercury bioaccumulates in water. Nearly every state has fish consumption advisories, especially for pregnant women, as fetuses are considered most vulnerable.

No federal rules on mercury from power plants are in place yet, although EPA determined in 2000 that regulation was "appropriate and necessary." Under existing law, there is only one way to regulate a hazardous air pollutant like mercury (as opposed to less dangerous pollutants). This so-called MACT (maximum achievable control technology) approach requires all polluters to meet an air standard based on the average emissions of the cleanest 12% of power plants.

While calculating the MACT, EPA became enamored of pollution-trading approaches, allowed by law for so-called criteria or conventional air pollutants. For instance, the "Clear Skies" legislation, introduced in Congress in June 2002, included a pollution-trading scheme to reduce emissions of sulfur dioxide (SO_2) and nitrogen oxides (NO_x). That's relevant to the mercury debate because the same scrubber technology that can clean up these pollutants can also reduce mercury in some situations, yielding what's called a "cobenefit."

After that bill stalled, EPA proposed a rule in January 2004 that would regulate mercury under a similar cap-and-trade system. The agency claimed that this trading approach would cut emissions by 70% to 15 tons by 2018—apparently a much better bottom line than the MACT approach, which EPA said would lower annual emissions to ▶

* Additional Analyses of Mercury Emissions Needed Before EPA Finalizes Rules for Coal-Fired Electric Utilities. www.epa.gov/oigearth/reports/2005/20050203-2005-P-00003-Gcopy.pdf

CREDIT: PHOTODISC RED/GETTY IMAGES

ScienceScope

EPA to Consider Human Pesticide Tests

The Environmental Protection Agency (EPA) will once again accept data from controversial studies that deliberately dose human volunteers with pesticides.

EPA stopped considering such data in December 2001, after the advocacy organization Environmental Working Group (EWG) challenged them as unethical. A review by the National Academy of Sciences (NAS) recommended that EPA accept the results of certain human tests if they met strict scientific and ethical criteria (*Science*, 27 February 2004, p. 1272). Meanwhile, CropLife America, a Washington, D.C.–based industry trade group, had sued EPA arguing that the moratorium was illegal, and in 2003 a judge agreed.

Now EPA has announced in an 8 February *Federal Register* notice that unless the studies are "fundamentally unethical," it will consider them case by case until new guidelines, including an ethics review board, are in place. That's consistent with the NAS recommendations. Still, EWG's Richard Wiles is upset. "This is the worst possible outcome," he says. "There are no rules, as far as I can tell."

–JOCELYN KAISER

Harvard Creates New Task Forces on Women in Science

A month after making controversial remarks about why men outnumber women in most scientific disciplines (*Science*, 28 January, p. 492), Harvard University president Lawrence Summers last week set up two task forces on campus to change the situation. The first, led by historian Evelyn Hammonds, will work to improve faculty searches and create a senior administrative position for improving gender diversity. The second group, chaired by computer scientist Barbara Grosz, will probe why women are underrepresented.

–YUDHIJIT BHATTACHARJEE

Nascent Reform Bill Criticized

PARIS—French scientists took to the streets last week to protest a government bill designed to boost research by reforming it (*Science*, 7 January, p. 27). The bill hasn't been made public yet, but after reviewing a leaked draft, leading scientists have concluded that it focuses too heavily on applied research. The government has scheduled more meetings with unions and leaders this month, so the bill won't be presented to Parliament until March at the earliest.

–BARBARA CASASSUS

Published by AAAS

only 34 tons by 2008. Industry likes this approach, because it gives power plants more flexibility in the technology they can employ and provides time to cope by slowly tightening the regulations.

Environmentalists and state regulatory agencies were highly critical, charging that the trading system would allow the dirtiest power plants to buy the rights to continue polluting, and mercury would continue to accumulate in toxic hot spots. In April of last year, seven senators asked the IG to investigate.

Now the IG has weighed in, charging in a 3 February report that EPA analyses were intentionally "biased" to make the MACT standard look less effective. Citing internal e-mails, the IG maintains that high-ranking officials had their fingers on the scale during this process: "EPA staff were instructed to develop a MACT standard that would result in national emissions of 34 tons per year" by 2008, the report found. Agency documents show that EPA took several stabs at running the model that produces the MACT standards, first yielding 29 tons, then 27, then finally 31. EPA then adjusted the results of the final run to hit the target.

Why 34 tons? The IG notes that's the same reduction that would be achieved as a cobenefit by simply reducing SO_2 and NO_x under the cap-and-trade rule proposed earlier. Martha Keating of the Clean Air Task Force in Boston, Massachusetts, sees it as an attempt to save industry from any extra costs. She says, and state regulators agree, that power plants could achieve greater reductions under MACT if they were required to install new technology, called activated carbon injection. EPA says it didn't generally consider the effects of this technology for its MACT standard, arguing that it won't be commercially ready by 2008.

The IG recommends that EPA rerun its analyses of the MACT standard and tighten its cap-and-trade proposal, but it can't force the agency to do so. EPA says that its final rule, expected 15 March, will include further details, analyses, and cost-benefit information. Spokesperson Cynthia Bergman maintains that the agency properly created the MACT standard and that the cap-and-trade rule is the better way to go. Meanwhile, Senator Jim Jeffords (D–VT), one of those who signed the request to the IG, called for "extensive oversight hearings into this important health issue and into the process by which this rule was crafted." **–ERIK STOKSTAD**

ASTRONOMY

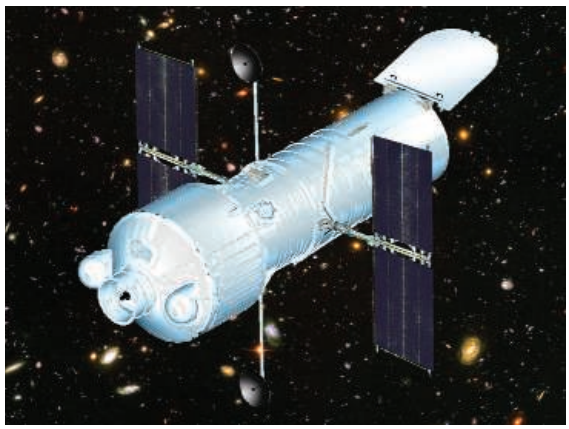
Hearing Highlights Dispute Over Hubble's Future

Scientists, engineers, and politicians are increasingly at odds over what to do with the Hubble Space Telescope. That much was clear at a contentious hearing last week before the House Science Committee, where participants disagreed over whether and how to service the aging spacecraft, what each option would cost, and how to pay for it.

Sean O'Keefe, set to give up his job as NASA administrator, caused a stir last year when he canceled a mission to have astronauts upgrade Hubble's instruments and keep it running until the end of the decade, when the James Webb Space Telescope is slated for launch. After pressure from lawmakers, he suggested that a robotic mission would be a safer bet than sending humans. That proposal, however, was shot down in December by a panel of the National Academy of Sciences, which called the robotic option too complex and costly and urged O'Keefe to reconsider sending astronauts to do the job. The panel also noted that the telescope could fail by 2007, before the robot likely would be ready. This week President George W. Bush requested no funding for a servicing mission in NASA's 2006 budget, a step that seems certain to keep the debate raging.

Representative Sherwood Boehlert (R–NY), who chairs the science committee, called himself an "agnostic" and pleaded with witnesses to "clarify what's at stake." What emerged were the deep divisions among scientists—including those at the same institution.

Astronomer Colin Norman of the Space Telescope Science Institute in Baltimore, Maryland, said the best option is to forgo fixing Hubble in favor of a $1 billion telescope, dubbed Hubble Origins Probe (HOP), that could examine dark energy, dark matter, and planets around other stars in addition to extending Hubble's mission. He noted that Japan has offered to help pay for HOP, which would be launched in 2010. "We must continue with the Hubble adventure," Norman added. The institute's director, Steven Beckwith, also favors completing Hubble's mission. But he wants to do it "as soon as possible," having it fixed by experienced astronauts aboard the shuttle rather than building and launching a new telescope. Other researchers expressed fear that any fix would come at the expense of other science projects.



Follow on. Instead of fixing Hubble, some astronomers are advocating a new telescope, the Hubble Origins Probe.

Joseph Taylor, a Princeton University astronomer who co-chaired the academy's 2000 astronomy study that set long-range priorities, says he opposes any servicing "if it requires major delays or reordering" of future missions. Neither a new telescope nor a servicing mission "should be a higher priority" than the Webb and Constellation-X, another planned NASA telescope, he stated.

Although astronomers are loath to lose Hubble, they also want to protect projects in the decadal study. "We have been playing fast and loose with the process by ignoring our prioritizations," says Alan Dressler of the Carnegie Institution of Washington in Pasadena, California, who did not testify at the hearing. Dressler wants the academy to find out which missions astronomers would be willing to sacrifice to save Hubble.

Louis Lanzerotti, who led the academy's Hubble study, agrees that if science must pay the servicing tab, priorities must be assessed. "If $1 billion is going to come out of some other aspect of NASA's science program, then I would have serious questions" about another Hubble mission—be it a new telescope, a shuttle service, or a robotic effort. But both he and Taylor would back a servicing mission if the money came from elsewhere. Lanzerotti added that NASA's $1 billion estimate doesn't square with the $300 million to $400 million price tag of past shuttle missions: "There is some accounting here which doesn't compute."

For many scientists, NASA's robotic mission is the least attractive option. Lanzerotti, for one, said it would be using an important scientific instrument as "target practice" for new technologies. But Representative Dana Rohrabacher (R–CA) argued that NASA should "push the envelope" by taking the opportunity to develop technologies that could benefit from Bush's plans for space exploration. **–ANDREW LAWLER**

With reporting by Robert Irion.

Published by AAAS

Many U.S. science agencies would have to make do with less under the president's 2006 budget request, which aims to cut the deficit, boost military and antiterrorism spending, and make tax cuts permanent

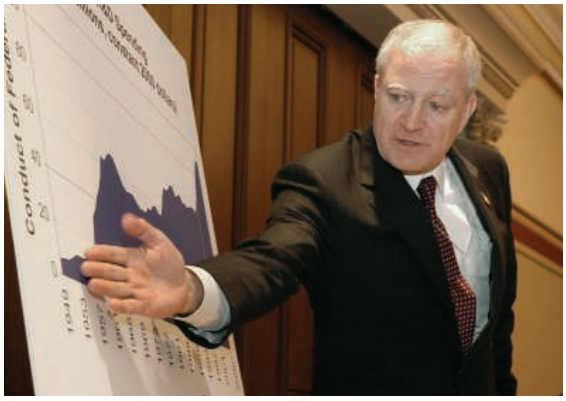
Caught in the Squeeze

President George W. Bush has proposed a flat budget for U.S. science next year. And the spinning has begun in earnest.

John Marburger, director of the White House Office of Science and Technology Policy, calls it "a pretty good year" for research, given the Administration's priorities of fighting terrorism, defending the homeland, and reducing the federal deficit. He says that the proposed 1% decline in the $61 billion federal science and technology budget for 2006—which excludes the Pentagon's even larger weapons development budget—would have been much worse but for the fact that "the president really cares about science."

However, most science policy analysts are wringing their hands over the tiny increase sought for the National Institutes of Health (NIH), a small rebound for the National Science Foundation (NSF) after a cut in 2005, and reductions in the science budgets at NASA, the National Oceanic and Atmospheric Administration, and the departments of energy and defense. At a time when other countries are ramping up their scientific efforts, they say, the United States shouldn't be resting on its laurels.

"The inadequate investments in research proposed by the Administration would erode the research and innovative capacity of our nation," says Nils Hasselmo, president of the 62-member Association of American Universities. AAU and the Federation of American Societies for Experimental Biology both call the president's 2006 request "disappointing," with FASEB adding that the proposed funding levels could "discourage our most talented young people from pursuing careers in biomedical research."



Taking the long view. Presidential science adviser John Marburger notes a sharp rise in science funding in Bush's first term.

The 0.7% increase for the $28 billion NIH, coming 2 years after a succession of double-digit boosts that resulted in a 5-year budget doubling, prompted agricultural imagery from newly installed Department of Health and Human Services (HHS) Secretary Michael Leavitt: "We have planted. It's now time for us to harvest the fruit." Even science-savvy legislators from the president's own party struggled to find a bright side. Representative Sherwood Boehlert (R–NY), chair of the House Science Committee, seized on an 8% boost to the $450 million intramural research budget at the National Institute of Standards and Technology (NIST) even as the president proposed eliminating a $137 million precompetitive technology research program the institute oversees. "Given an overall cut to nondefense domestic discretionary spending, science programs fared relatively well," Boehlert noted. "I was especially pleased to see the significant increases for the NIST labs."

The 2006 budget request, following tradition, unfolded in a series of briefings by agency heads. Here are some highlights, brought to you by *Science* reporters who were there.

NIH: The president's budget includes a 42% boost, to $333 million, for a set of cross-NIH initiatives to support translational research, known collectively as the Roadmap. Biodefense efforts would receive a 3.2% hike, to $18 billion, and another $26 million would be allocated for the Neuroscience Blueprint involving 15 institutes. NIH is also getting $97 million more to develop countermeasures for a radiological or chemical attack.

Still, the overall news is grim, as most institutes would get increases averaging about 0.5%. NIH Director Elias Zerhouni says he hopes to protect the number of funded investigators by shifting money from some clinical and center grants that expire in 2006 into new and competing grants, which will rise for the first time since 2003. But the average grant size of $347,000 would remain flat, and the proportion of applications funded would continue to

Science Education Takes a Hit at NSF

The National Science Foundation's (NSF's) role in improving science and math education in the United States would shrink significantly under the president's 2006 budget request. Particularly hard hit are programs to improve the skills of elementary and secondary school science and mathematics teachers, develop new teaching materials, and evaluate whether those activities are working.

"This is outrageous," says Gerry Wheeler, executive director of the National Science Teachers Association. "Despite all the concern about how U.S. students perform on international math and science tests, the Administration has made it clear that K–12 science education is not a priority."

The request would trim the budget for NSF's Education and Human Resources (EHR) directorate by $104 million, to $737 million, a 12.4% drop that follows a similar reduction this year. By NSF's own estimate, its programs would reach 64,000 elementary and secondary school students and teachers in 2006, compared with 100,000 in 2004.

The biggest blow would fall on the directorate's division of elementary, secondary, and informal education. A $60 million program begun last year to help teachers, from training them to providing professional development, would be slashed by nearly half, to $33 million. A $28 million program to develop new classroom materials and focus on an increasingly diverse student population would be pared by one-third, and a university-based network of Centers for Learning and Teaching, with 16 sites, would make no new awards in

CREDIT: MARTY KATZ

Published by AAAS

plummet, to a projected 21%. NIH is boosting postdoc stipends by 4% and increasing health benefits. But the result is a 2% drop in the number that would be supported. "We think it's the right choice," Zerhouni says.

NSF: A $113 million increase proposed for the agency's $4.2 billion research budget hides a $48 million transfer from the U.S. Coast Guard to take on the annual cost of breaking ice to keep the shipping lanes open in the Antarctic. The 2006 request includes funding for all five of the agency's major new facilities under construction, but it lacks two expected new starts in 2006: a network of ocean observatories and an Alaskan regional research vessel. NSF Director Arden Bement says he hopes to request money for them in 2007 if the budget climate warms up. The biggest hit comes in the agency's education programs (see sidebar below).

NASA: The news was good for missions that would support the president's vision for eventual lunar and martian exploration by humans. The lunar program, which would be focused on technology more than science, would nearly triple, from $52 million to $135 million, and Mars projects would jump from $681 million to $723 million. The largest increase would go to developing a rocket capable of taking humans beyond Earth orbit; the Constellation project would more than double, to $1.12 billion in 2006.The one exception to that rule is human research: Funding for studying the effects of space on astronauts would plummet from $1 billion to $807 million for 2006.

The biggest losers would be missions to the outer planets and Earth-observing activities (see sidebar on p. 834).

Energy: As part of the 4% cut for the Office of Science, Department of Energy officials want to pull the plug on a $140 million experiment at Fermi National Accelerator Laboratory (Fermilab) in Batavia, Illinois, to study the physics of particles that contain the bottom quark. Science chief Ray Orbach says the Large Hadron Collider being completed at CERN, the European particle physics lab near Geneva, Switzerland, would cover the same territory as BTeV, which was set to begin construction this year, and that the savings will go toward developing a future neutrino detector at Fermilab. "Maybe it's not that they're trying to drive science from the United States, but boy, they're sure making it look like they are," says Sheldon Stone, a physicist at Syracuse University and BTeV co-spokesperson. Operations at the Relativistic Heavy Ion Collider, the primary accelerator at Brookhaven National Laboratory in Upton, New York, will be curtailed, with funding for only 1400 hours of experiments compared with a scheduled 3600 hours this year.

FDA: The Food and Drug Administration wants $30 million more to expand a network of state labs that can handle threats to food safety, an area former HHS secretary Tommy Thompson says is vulnerable to terrorism. It also hopes to hire 25 more people to clear up a backlog of reports submitted on potential safety problems with drugs that are already on the market. "It's a step in the right direction," says Jerry Avorn, a pharmacoepidemiologist at Harvard Medical School in Boston

2006 U.S. Budget Highlights (In $ Millions)

	2005	2006 request	Annual % change
NIH	28,650	28,845	0.7%
NSF	5,472	5,605	2.4%
Research	4,220	4,333	2.7%
Education	841	737	–12.4%
Major facilities	174	250	43.7%
NASA	16,070	16,456	2.4%
Science	5,530	5,480	–0.9%
Aeronautics	906	852	–6.0%
Department of Defense basic research	1,513	1319	–12.8%
Department of Energy Office of Science	3,610	3463	–4.1%
High-energy physics	736	714	–3.0%
Basic energy sciences	1105	1146	3.7%
Nuclear physics	405	371	–8.4%
Department of Homeland Security science	1,115	1368	22.7%
Department of Commerce			
NOAA research	579	527	–8.4%
NIST science and technology research	380	418	10.0%
NIST Advanced Technology Program	136	(cut)	
Environmental Protection Agency R&D	744	760	2.2%
Geological Survey	935	934	–0.1%
USDA National Research Initiative	180	250	38.9%
FDA	1801	1881	4.4%
Agricultural Research Service	1306	1079	–17.4%
Total R&D*	**131,571**	**132,304**	**0.6%**

** Includes agencies not listed.*

TABLE SOURCE: U.S. GOVERNMENT AGENCIES; PHOTO CREDIT: GETTY IMAGES

Left behind. NSF's programs will reach 36,000 fewer students in 2006 than in 2004.

2006. In addition, the math and science partnerships program, begun in 2002 as a $200-million-a-year effort to link university science and engineering departments with their local school districts, would continue to wind down, with only enough money to fulfill existing commitments.

The biggest percentage loser in the 2006 budget is the directorate's $59 million division of research evaluation, targeted for a 43% drop. NSF officials project that the president's request will mean no new awards next year for programs aimed at developing new ways to monitor the performance of students and teachers as well as evaluating the effectiveness of new methods and materials.

NSF Director Arden Bement says that the EHR reductions give NSF the chance "to sharpen our focus on programs with a proven track record. ... We have a lot of knowledge of what needs to be done. Now we have to do it."

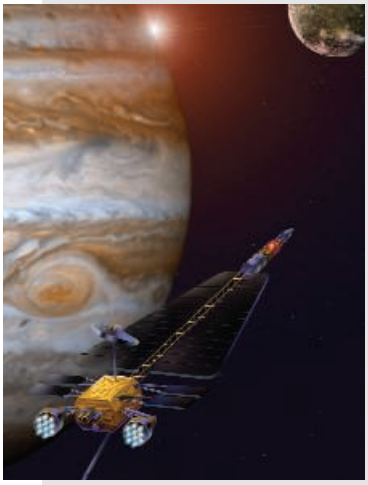
–JEFFREY MERVIS

Published by AAAS

Other Highlights From the Budget

Jupiter Is a Blue State, Mars Is Red

The timing could not be more ironic. Just as a joint U.S.-European spacecraft is making exciting and front-page discoveries from distant Saturn, the White House proposes a budget that could scrub the agency's only major mission planned for the outer solar system. Another victim is an earth science flight to study aerosols, and several other longer-term projects, from planet finders to dark-energy seekers, would be put on hold.



Slipping. The Jupiter Icy Moons Orbiter would be delayed at least 6 years.

The strains placed on NASA by the Columbia failure and U.S. President George W. Bush's new exploration vision are evident in the request, which includes only half the increase the White House had promised last year for 2006. Outgoing NASA Administrator Sean O'Keefe says the request would have been far worse without the exploration plan Bush laid out last January: "It's rather remarkable, given the circumstances."

The request would not cut any "ongoing" science programs, says science chief Al Diaz, whose budget would stay relatively flat. But a host of missions still in the early stages of planning would be delayed, some indefinitely. The most dramatic impact would be on the Jupiter Icy Moons Orbiter (JIMO), an elaborate and expensive mission that would harness nuclear electric technology to provide unprecedented access to Europa and the giant planet's other moons. The technology made JIMO "a high-risk venture," says Craig Steidle, NASA exploration chief. Technology funding for the mission would be slashed from $432 million to $320 million, and JIMO would be delayed at least until 2018—6 years later than NASA officials had projected just a year ago.

Instead, Diaz said NASA would reconsider a simpler mission to Europa that was canceled in 2002. Diaz says it may be included in a revamped science strategy this summer.

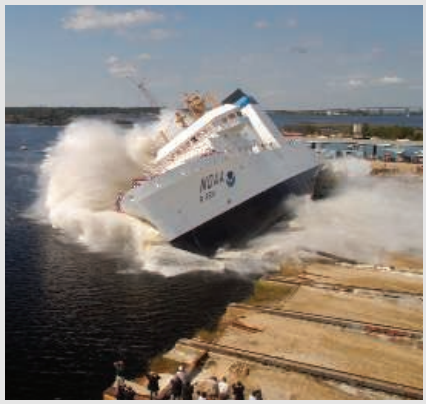
The request contains bad news for scientists working in other fields. The launch date for the Kepler mission, designed to search for extrasolar planets, has slipped from 2007 to "to be determined," according to NASA documents. The Dawn project, which would visit the asteroid belt, would be downsized and delayed. And the 2007 flight of the Glory satellite, which would measure atmospheric aerosols, would be abandoned, although some of its instruments might be used on other spacecraft. Technical challenges will delay the Space Interferometry Mission, planned for launch in the next decade to search for Earth-sized planets. And the Beyond Einstein program, which would launch a series of spacecraft to test Einstein's theories (p. 869), remains a dream. **–ANDREW LAWLER**

Ocean Research Budget Ebbs

Ocean policy is hot, but advocates say that President George W. Bush's proposed budget is tepid when it comes to addressing the needs of the nation's troubled waters. A 10% cut in the $580 million research budget for the National Oceanic and Atmospheric Administration (NOAA), the government's key ocean research and protection agency, "provides a rather distressing signal about the level of commitment [to the oceans]," says Ted Morton, federal policy director for a Washington, D.C.–based advocacy group called Oceana.

Not so, says NOAA Deputy Administrator James Mahoney. The 2005 figure was inflated by legislative earmarks, he says. A more accurate measure of the Administration's commitment, he argues, is that the president requested 7% more for NOAA than he asked for last year.

Last fall a presidential commission urged the White House to devote more attention to the Great Lakes and coastal and marine resources and said $1.5 billion was needed to jump-start a successful national ocean program. Three months later, the president's U.S. Action Ocean Plan established a Cabinet-level, interagency task force on oceans (*Science*, 24 December 2004, p. 2171). The 2006 request is the next step, says Mahoney.



Full steam ahead. A new NOAA fisheries survey ship is in the works.

While some NOAA programs are being squeezed, a few efforts tied to marine research are getting boosts. The agency has requested $38.5 million for a new fisheries survey vessel, $1.5 million more for the $25 million coral reef program, and $10 million for an expanded tsunami warning system (*Science*, 21 January, p. 331). In a surprise move, the White House submitted a level budget for Sea Grant, which supports marine and Great Lakes research and education in coastal states. The program historically has relied on Congress to keep it healthy.

–ELIZABETH PENNISI

and the author of *Powerful Medicines: The Benefits, Risks, and Costs of Prescription Drugs*. But any changes, he says, also require a new "culture of openness."

Homeland Security: The department wants $227 million for a new Domestic Nuclear Detection Office to sniff out attempts to bring bombs into the country. Several federal agencies will contribute staffers to the new office, which President Bush mentioned in last month's State of the Union address.

Defense: Although the Pentagon's basic research account would slump by 13%, officials hope to scale up a pilot scholarship program to attract more U.S. citizens into government defense jobs. The first 25 awards in the Science, Mathematics, and Research for Transformation program are due to be announced this spring, and the 2006 request would allow for up to 100 2-year undergraduate and graduate scholarships in 15 disciplines.

Graduates must return the favor by working for the department. But Michael Corradini, a mechanical engineer at the University of Wisconsin, Madison, worries that the requirement could scare off potential applicants. He suggests instead that graduates should be required to do a summer internship in the department. "If students have a meaningful experience during the internships," he says, "they might be inclined to pursue a DOD career."

The $2.5 trillion proposed budget now goes to Congress, which will tinker with the president's priorities and add in its own. That means the fate of these and other research programs, although traditionally nonpartisan, will be shaped by larger forces—from Social Security to tax policy—stirring the political waters.

–JEFFREY MERVIS

With reporting by Amitabh Avasthi, Yudhijit Bhattacharjee, Jennifer Couzin, Marie Granmar, Jocelyn Kaiser, Eli Kintisch, Andrew Lawler, and Charles Seife.

CREDITS (TOP TO BOTTOM): NASA/JPL; NOAA

Published by AAAS

Many U.S. science agencies would have to make do with less under the president's 2006 budget request, which aims to cut the deficit, boost military and antiterrorism spending, and make tax cuts permanent

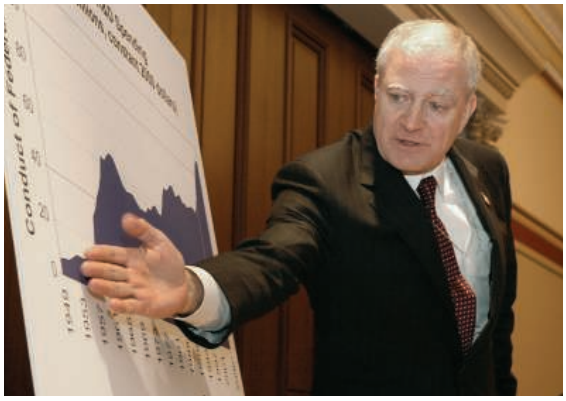
Caught in the Squeeze

President George W. Bush has proposed a flat budget for U.S. science next year. And the spinning has begun in earnest.

John Marburger, director of the White House Office of Science and Technology Policy, calls it "a pretty good year" for research, given the Administration's priorities of fighting terrorism, defending the homeland, and reducing the federal deficit. He says that the proposed 1% decline in the $61 billion federal science and technology budget for 2006—which excludes the Pentagon's even larger weapons development budget—would have been much worse but for the fact that "the president really cares about science."

However, most science policy analysts are wringing their hands over the tiny increase sought for the National Institutes of Health (NIH), a small rebound for the National Science Foundation (NSF) after a cut in 2005, and reductions in the science budgets at NASA, the National Oceanic and Atmospheric Administration, and the departments of energy and defense. At a time when other countries are ramping up their scientific efforts, they say, the United States shouldn't be resting on its laurels.

"The inadequate investments in research proposed by the Administration would erode the research and innovative capacity of our nation," says Nils Hasselmo, president of the 62-member Association of American Universities. AAU and the Federation of American Societies for Experimental Biology both call the president's 2006 request "disappointing," with FASEB adding that the proposed funding levels could "discourage our most talented young people from pursuing careers in biomedical research."



Taking the long view. Presidential science adviser John Marburger notes a sharp rise in science funding in Bush's first term.

The 0.7% increase for the $28 billion NIH, coming 2 years after a succession of double-digit boosts that resulted in a 5-year budget doubling, prompted agricultural imagery from newly installed Department of Health and Human Services (HHS) Secretary Michael Leavitt: "We have planted. It's now time for us to harvest the fruit." Even science-savvy legislators from the president's own party struggled to find a bright side. Representative Sherwood Boehlert (R–NY), chair of the House Science Committee, seized on an 8% boost to the $450 million intramural research budget at the National Institute of Standards and Technology (NIST) even as the president proposed eliminating a $137 million precompetitive technology research program the institute oversees. "Given an overall cut to nondefense domestic discretionary spending, science programs fared relatively well," Boehlert noted. "I was especially pleased to see the significant increases for the NIST labs."

The 2006 budget request, following tradition, unfolded in a series of briefings by agency heads. Here are some highlights, brought to you by *Science* reporters who were there.

NIH: The president's budget includes a 42% boost, to $333 million, for a set of cross-NIH initiatives to support translational research, known collectively as the Roadmap. Biodefense efforts would receive a 3.2% hike, to $18 billion, and another $26 million would be allocated for the Neuroscience Blueprint involving 15 institutes. NIH is also getting $97 million more to develop countermeasures for a radiological or chemical attack.

Still, the overall news is grim, as most institutes would get increases averaging about 0.5%. NIH Director Elias Zerhouni says he hopes to protect the number of funded investigators by shifting money from some clinical and center grants that expire in 2006 into new and competing grants, which will rise for the first time since 2003. But the average grant size of $347,000 would remain flat, and the proportion of applications funded would continue to

Science Education Takes a Hit at NSF

The National Science Foundation's (NSF's) role in improving science and math education in the United States would shrink significantly under the president's 2006 budget request. Particularly hard hit are programs to improve the skills of elementary and secondary school science and mathematics teachers, develop new teaching materials, and evaluate whether those activities are working.

"This is outrageous," says Gerry Wheeler, executive director of the National Science Teachers Association. "Despite all the concern about how U.S. students perform on international math and science tests, the Administration has made it clear that K–12 science education is not a priority."

The request would trim the budget for NSF's Education and Human Resources (EHR) directorate by $104 million, to $737 million, a 12.4% drop that follows a similar reduction this year. By NSF's own estimate, its programs would reach 64,000 elementary and secondary school students and teachers in 2006, compared with 100,000 in 2004.

The biggest blow would fall on the directorate's division of elementary, secondary, and informal education. A $60 million program begun last year to help teachers, from training them to providing professional development, would be slashed by nearly half, to $33 million. A $28 million program to develop new classroom materials and focus on an increasingly diverse student population would be pared by one-third, and a university-based network of Centers for Learning and Teaching, with 16 sites, would make no new awards in

CREDIT: MARTY KATZ

Published by AAAS

plummet, to a projected 21%. NIH is boosting postdoc stipends by 4% and increasing health benefits. But the result is a 2% drop in the number that would be supported. "We think it's the right choice," Zerhouni says.

NSF: A $113 million increase proposed for the agency's $4.2 billion research budget hides a $48 million transfer from the U.S. Coast Guard to take on the annual cost of breaking ice to keep the shipping lanes open in the Antarctic. The 2006 request includes funding for all five of the agency's major new facilities under construction, but it lacks two expected new starts in 2006: a network of ocean observatories and an Alaskan regional research vessel. NSF Director Arden Bement says he hopes to request money for them in 2007 if the budget climate warms up. The biggest hit comes in the agency's education programs (see sidebar below).

NASA: The news was good for missions that would support the president's vision for eventual lunar and martian exploration by humans. The lunar program, which would be focused on technology more than science, would nearly triple, from $52 million to $135 million, and Mars projects would jump from $681 million to $723 million. The largest increase would go to developing a rocket capable of taking humans beyond Earth orbit; the Constellation project would more than double, to $1.12 billion in 2006.The one exception to that rule is human research: Funding for studying the effects of space on astronauts would plummet from $1 billion to $807 million for 2006.

The biggest losers would be missions to the outer planets and Earth-observing activities (see sidebar on p. 834).

Energy: As part of the 4% cut for the Office of Science, Department of Energy officials want to pull the plug on a $140 million experiment at Fermi National Accelerator Laboratory (Fermilab) in Batavia, Illinois, to study the physics of particles that contain the bottom quark. Science chief Ray Orbach says the Large Hadron Collider being completed at CERN, the European particle physics lab near Geneva, Switzerland, would cover the same territory as BTeV, which was set to begin construction this year, and that the savings will go toward developing a future neutrino detector at Fermilab. "Maybe it's not that they're trying to drive science from the United States, but boy, they're sure making it look like they are," says Sheldon Stone, a physicist at Syracuse University and BTeV co-spokesperson. Operations at the Relativistic Heavy Ion Collider, the primary accelerator at Brookhaven National Laboratory in Upton, New York, will be curtailed, with funding for only 1400 hours of experiments compared with a scheduled 3600 hours this year.

FDA: The Food and Drug Administration wants $30 million more to expand a network of state labs that can handle threats to food safety, an area former HHS secretary Tommy Thompson says is vulnerable to terrorism. It also hopes to hire 25 more people to clear up a backlog of reports submitted on potential safety problems with drugs that are already on the market. "It's a step in the right direction," says Jerry Avorn, a pharmacoepidemiologist at Harvard Medical School in Boston

2006 U.S. Budget Highlights (In $ Millions)

	2005	2006 request	Annual % change
NIH	28,650	28,845	0.7%
NSF	5,472	5,605	2.4%
Research	4,220	4,333	2.7%
Education	841	737	–12.4%
Major facilities	174	250	43.7%
NASA	16,070	16,456	2.4%
Science	5,530	5,480	–0.9%
Aeronautics	906	852	–6.0%
Department of Defense basic research	1,513	1319	–12.8%
Department of Energy Office of Science	3,610	3463	–4.1%
High-energy physics	736	714	–3.0%
Basic energy sciences	1105	1146	3.7%
Nuclear physics	405	371	–8.4%
Department of Homeland Security science	1,115	1368	22.7%
Department of Commerce			
NOAA research	579	527	–8.4%
NIST science and technology research	380	418	10.0%
NIST Advanced Technology Program	136	(cut)	
Environmental Protection Agency R&D	744	760	2.2%
Geological Survey	935	934	–0.1%
USDA National Research Initiative	180	250	38.9%
FDA	1801	1881	4.4%
Agricultural Research Service	1306	1079	–17.4%
Total R&D*	**131,571**	**132,304**	**0.6%**

** Includes agencies not listed.*

TABLE SOURCE: U.S. GOVERNMENT AGENCIES; PHOTO CREDIT: GETTY IMAGES

Left behind. NSF's programs will reach 36,000 fewer students in 2006 than in 2004.

2006. In addition, the math and science partnerships program, begun in 2002 as a $200-million-a-year effort to link university science and engineering departments with their local school districts, would continue to wind down, with only enough money to fulfill existing commitments.

The biggest percentage loser in the 2006 budget is the directorate's $59 million division of research evaluation, targeted for a 43% drop. NSF officials project that the president's request will mean no new awards next year for programs aimed at developing new ways to monitor the performance of students and teachers as well as evaluating the effectiveness of new methods and materials.

NSF Director Arden Bement says that the EHR reductions give NSF the chance "to sharpen our focus on programs with a proven track record. ... We have a lot of knowledge of what needs to be done. Now we have to do it."

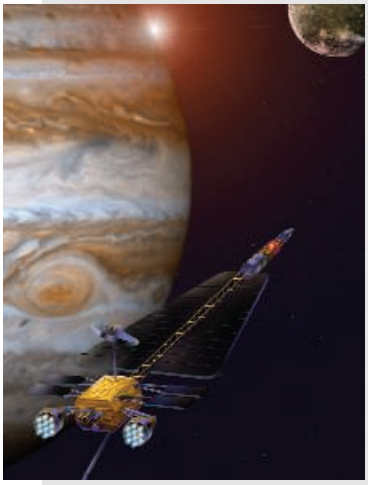
–JEFFREY MERVIS

Published by AAAS

Other Highlights From the Budget

Jupiter Is a Blue State, Mars Is Red

The timing could not be more ironic. Just as a joint U.S.-European spacecraft is making exciting and front-page discoveries from distant Saturn, the White House proposes a budget that could scrub the agency's only major mission planned for the outer solar system. Another victim is an earth science flight to study aerosols, and several other longer-term projects, from planet finders to dark-energy seekers, would be put on hold.



Slipping. The Jupiter Icy Moons Orbiter would be delayed at least 6 years.

The strains placed on NASA by the Columbia failure and U.S. President George W. Bush's new exploration vision are evident in the request, which includes only half the increase the White House had promised last year for 2006. Outgoing NASA Administrator Sean O'Keefe says the request would have been far worse without the exploration plan Bush laid out last January: "It's rather remarkable, given the circumstances."

The request would not cut any "ongoing" science programs, says science chief Al Diaz, whose budget would stay relatively flat. But a host of missions still in the early stages of planning would be delayed, some indefinitely. The most dramatic impact would be on the Jupiter Icy Moons Orbiter (JIMO), an elaborate and expensive mission that would harness nuclear electric technology to provide unprecedented access to Europa and the giant planet's other moons. The technology made JIMO "a high-risk venture," says Craig Steidle, NASA exploration chief. Technology funding for the mission would be slashed from $432 million to $320 million, and JIMO would be delayed at least until 2018—6 years later than NASA officials had projected just a year ago.

Instead, Diaz said NASA would reconsider a simpler mission to Europa that was canceled in 2002. Diaz says it may be included in a revamped science strategy this summer.

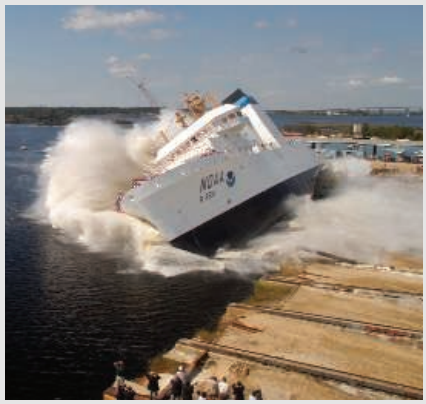
The request contains bad news for scientists working in other fields. The launch date for the Kepler mission, designed to search for extrasolar planets, has slipped from 2007 to "to be determined," according to NASA documents. The Dawn project, which would visit the asteroid belt, would be downsized and delayed. And the 2007 flight of the Glory satellite, which would measure atmospheric aerosols, would be abandoned, although some of its instruments might be used on other spacecraft. Technical challenges will delay the Space Interferometry Mission, planned for launch in the next decade to search for Earth-sized planets. And the Beyond Einstein program, which would launch a series of spacecraft to test Einstein's theories (p. 869), remains a dream. **–ANDREW LAWLER**

Ocean Research Budget Ebbs

Ocean policy is hot, but advocates say that President George W. Bush's proposed budget is tepid when it comes to addressing the needs of the nation's troubled waters. A 10% cut in the $580 million research budget for the National Oceanic and Atmospheric Administration (NOAA), the government's key ocean research and protection agency, "provides a rather distressing signal about the level of commitment [to the oceans]," says Ted Morton, federal policy director for a Washington, D.C.–based advocacy group called Oceana.

Not so, says NOAA Deputy Administrator James Mahoney. The 2005 figure was inflated by legislative earmarks, he says. A more accurate measure of the Administration's commitment, he argues, is that the president requested 7% more for NOAA than he asked for last year.

Last fall a presidential commission urged the White House to devote more attention to the Great Lakes and coastal and marine resources and said $1.5 billion was needed to jump-start a successful national ocean program. Three months later, the president's U.S. Action Ocean Plan established a Cabinet-level, interagency task force on oceans (*Science*, 24 December 2004, p. 2171). The 2006 request is the next step, says Mahoney.



Full steam ahead. A new NOAA fisheries survey ship is in the works.

While some NOAA programs are being squeezed, a few efforts tied to marine research are getting boosts. The agency has requested $38.5 million for a new fisheries survey vessel, $1.5 million more for the $25 million coral reef program, and $10 million for an expanded tsunami warning system (*Science*, 21 January, p. 331). In a surprise move, the White House submitted a level budget for Sea Grant, which supports marine and Great Lakes research and education in coastal states. The program historically has relied on Congress to keep it healthy.

–ELIZABETH PENNISI

and the author of *Powerful Medicines: The Benefits, Risks, and Costs of Prescription Drugs*. But any changes, he says, also require a new "culture of openness."

Homeland Security: The department wants $227 million for a new Domestic Nuclear Detection Office to sniff out attempts to bring bombs into the country. Several federal agencies will contribute staffers to the new office, which President Bush mentioned in last month's State of the Union address.

Defense: Although the Pentagon's basic research account would slump by 13%, officials hope to scale up a pilot scholarship program to attract more U.S. citizens into government defense jobs. The first 25 awards in the Science, Mathematics, and Research for Transformation program are due to be announced this spring, and the 2006 request would allow for up to 100 2-year undergraduate and graduate scholarships in 15 disciplines.

Graduates must return the favor by working for the department. But Michael Corradini, a mechanical engineer at the University of Wisconsin, Madison, worries that the requirement could scare off potential applicants. He suggests instead that graduates should be required to do a summer internship in the department. "If students have a meaningful experience during the internships," he says, "they might be inclined to pursue a DOD career."

The $2.5 trillion proposed budget now goes to Congress, which will tinker with the president's priorities and add in its own. That means the fate of these and other research programs, although traditionally nonpartisan, will be shaped by larger forces—from Social Security to tax policy—stirring the political waters.

–JEFFREY MERVIS

With reporting by Amitabh Avasthi, Yudhijit Bhattacharjee, Jennifer Couzin, Marie Granmar, Jocelyn Kaiser, Eli Kintisch, Andrew Lawler, and Charles Seife.

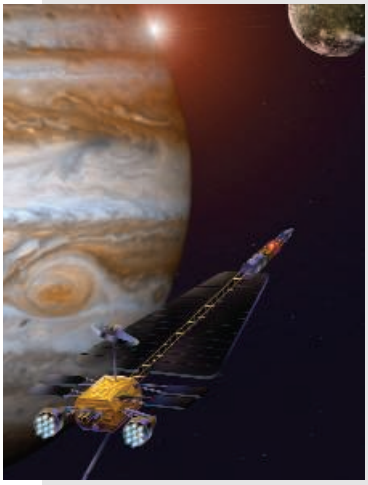
CREDITS (TOP TO BOTTOM): NASA/JPL; NOAA

Published by AAAS

Other Highlights From the Budget

Jupiter Is a Blue State, Mars Is Red

The timing could not be more ironic. Just as a joint U.S.-European spacecraft is making exciting and front-page discoveries from distant Saturn, the White House proposes a budget that could scrub the agency's only major mission planned for the outer solar system. Another victim is an earth science flight to study aerosols, and several other longer-term projects, from planet finders to dark-energy seekers, would be put on hold.



Slipping. The Jupiter Icy Moons Orbiter would be delayed at least 6 years.

The strains placed on NASA by the Columbia failure and U.S. President George W. Bush's new exploration vision are evident in the request, which includes only half the increase the White House had promised last year for 2006. Outgoing NASA Administrator Sean O'Keefe says the request would have been far worse without the exploration plan Bush laid out last January: "It's rather remarkable, given the circumstances."

The request would not cut any "ongoing" science programs, says science chief Al Diaz, whose budget would stay relatively flat. But a host of missions still in the early stages of planning would be delayed, some indefinitely. The most dramatic impact would be on the Jupiter Icy Moons Orbiter (JIMO), an elaborate and expensive mission that would harness nuclear electric technology to provide unprecedented access to Europa and the giant planet's other moons. The technology made JIMO "a high-risk venture," says Craig Steidle, NASA exploration chief. Technology funding for the mission would be slashed from $432 million to $320 million, and JIMO would be delayed at least until 2018—6 years later than NASA officials had projected just a year ago.

Instead, Diaz said NASA would reconsider a simpler mission to Europa that was canceled in 2002. Diaz says it may be included in a revamped science strategy this summer.

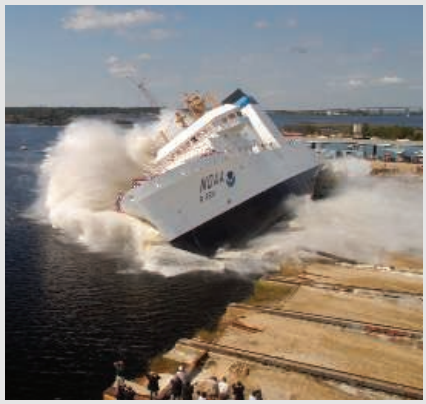
The request contains bad news for scientists working in other fields. The launch date for the Kepler mission, designed to search for extrasolar planets, has slipped from 2007 to "to be determined," according to NASA documents. The Dawn project, which would visit the asteroid belt, would be downsized and delayed. And the 2007 flight of the Glory satellite, which would measure atmospheric aerosols, would be abandoned, although some of its instruments might be used on other spacecraft. Technical challenges will delay the Space Interferometry Mission, planned for launch in the next decade to search for Earth-sized planets. And the Beyond Einstein program, which would launch a series of spacecraft to test Einstein's theories (p. 869), remains a dream. **–ANDREW LAWLER**

Ocean Research Budget Ebbs

Ocean policy is hot, but advocates say that President George W. Bush's proposed budget is tepid when it comes to addressing the needs of the nation's troubled waters. A 10% cut in the $580 million research budget for the National Oceanic and Atmospheric Administration (NOAA), the government's key ocean research and protection agency, "provides a rather distressing signal about the level of commitment [to the oceans]," says Ted Morton, federal policy director for a Washington, D.C.–based advocacy group called Oceana.

Not so, says NOAA Deputy Administrator James Mahoney. The 2005 figure was inflated by legislative earmarks, he says. A more accurate measure of the Administration's commitment, he argues, is that the president requested 7% more for NOAA than he asked for last year.

Last fall a presidential commission urged the White House to devote more attention to the Great Lakes and coastal and marine resources and said $1.5 billion was needed to jump-start a successful national ocean program. Three months later, the president's U.S. Action Ocean Plan established a Cabinet-level, interagency task force on oceans (*Science*, 24 December 2004, p. 2171). The 2006 request is the next step, says Mahoney.



Full steam ahead. A new NOAA fisheries survey ship is in the works.

While some NOAA programs are being squeezed, a few efforts tied to marine research are getting boosts. The agency has requested $38.5 million for a new fisheries survey vessel, $1.5 million more for the $25 million coral reef program, and $10 million for an expanded tsunami warning system (*Science*, 21 January, p. 331). In a surprise move, the White House submitted a level budget for Sea Grant, which supports marine and Great Lakes research and education in coastal states. The program historically has relied on Congress to keep it healthy.

–ELIZABETH PENNISI

and the author of *Powerful Medicines: The Benefits, Risks, and Costs of Prescription Drugs*. But any changes, he says, also require a new "culture of openness."

Homeland Security: The department wants $227 million for a new Domestic Nuclear Detection Office to sniff out attempts to bring bombs into the country. Several federal agencies will contribute staffers to the new office, which President Bush mentioned in last month's State of the Union address.

Defense: Although the Pentagon's basic research account would slump by 13%, officials hope to scale up a pilot scholarship program to attract more U.S. citizens into government defense jobs. The first 25 awards in the Science, Mathematics, and Research for Transformation program are due to be announced this spring, and the 2006 request would allow for up to 100 2-year undergraduate and graduate scholarships in 15 disciplines.

Graduates must return the favor by working for the department. But Michael Corradini, a mechanical engineer at the University of Wisconsin, Madison, worries that the requirement could scare off potential applicants. He suggests instead that graduates should be required to do a summer internship in the department. "If students have a meaningful experience during the internships," he says, "they might be inclined to pursue a DOD career."

The $2.5 trillion proposed budget now goes to Congress, which will tinker with the president's priorities and add in its own. That means the fate of these and other research programs, although traditionally nonpartisan, will be shaped by larger forces—from Social Security to tax policy—stirring the political waters.

–JEFFREY MERVIS

With reporting by Amitabh Avasthi, Yudhijit Bhattacharjee, Jennifer Couzin, Marie Granmar, Jocelyn Kaiser, Eli Kintisch, Andrew Lawler, and Charles Seife.

CREDITS (TOP TO BOTTOM): NASA/JPL; NOAA

Published by AAAS

Lupus Drug Company Asks FDA For Second Chance

Biotech firm pleads for drug's approval so it can afford to prove that it works

Doctors who treat the autoimmune disease lupus are on edge as the U.S. Food and Drug Administration (FDA) considers a rare plea: Approve a lupus drug that the agency has already rejected and that—even its maker acknowledges—has uncertain efficacy. The small California company that's developing the drug, La Jolla Pharmaceutical Company (LJPC), says it can't afford to complete the new clinical trial FDA has requested—the company's 14th, which it has already begun at a cost of $2.5 million a month. After meetings of lupus specialists, company executives, and FDA officials last fall, the agency is considering whether to approve the drug on the condition that LJPC conduct a postmarketing study to determine whether it works. It may rule this month.

The lupus community is split over whether the drug, LJP 394, should become the first therapy approved for the condition in over 30 years. The drug has an outstanding safety record, and some of the more than 500 lupus patients who've tried the therapy suffered fewer kidney flares, a hallmark of serious disease. Still, the trials sponsored by LJPC so far have failed to show definitively that it works.

"What you're left with is this terrible dilemma," says David Wofsy, a lupus specialist at the University of California, San Francisco, who was not involved in developing LJP 394. "There's an important unanswered question here, and it should be answered. But it's different than saying the drug should be approved."

LJPC has already spent close to $300 million—90% to 95% of its expenditures—on LJP 394, according to the company's chair and CEO, Steven Engle. Last week it raised $16 million, enough to see it through this year—although not enough to complete additional testing. Approval, Engle hopes, could bring not only revenue but also a corporate partner.

Doctors and patients are desperate for any new lupus drug because current therapy is so inadequate. Just three drugs—the steroid prednisone, the chemotherapy drug cyclophosphamide, and aspirin—are approved for the disease, which can attack nearly any organ.

Fifteen years ago, LJPC set out to change that. The company had patented a technology that disables a narrow swath of immune cells: B cells sporting anti-DNA antibodies. Such antibodies are common, although not universal, in the blood of lupus patients. They also appear in the kidneys of those with lupus-induced kidney disease, which strikes about a fifth of sufferers. Furthermore, several studies showed that a rise in anti-DNA antibody levels presages a kidney flare.

Tackling a butterfly. It's uncertain whether a new drug can erase the symptoms of lupus, often characterized by a "butterfly" rash.

In a phase II/III trial of LJP 394 in the late 1990s with 200 volunteers, LJPC teamed up with the pharmaceutical giant Abbott Laboratories, based in Abbott Park, Illinois. But in 1999, before the trial ended, Abbott pulled out. Abbott was concerned that the drug was ineffective, according to Engle, but an Abbott spokesperson says the company simply decided not to pursue treatments for lupus nephritis.

Does It Work?

	Phase III participants	Renal flares	Major flares in all organs	Needed high-dose immunosuppressants
LJP 394	145	17 (12%)	35 (24%)	32 (22%)
Placebo	153	24 (16%)	47 (31%)	36 (24%)

The drug failed when all the trial's subjects were considered. But when LJPC took a closer look at the data, it found that roughly 90% of patients in the trial had "high affinity" antibodies, to which the drug was likelier to bind, and those patients seemed to fare better than the rest. LJPC then forged ahead on its own with a phase III study that focused primarily on how those with high-affinity antibodies responded to the drug.

The results, announced in early 2003, were not what the drug's enthusiasts had hoped for. Twelve percent of kidney flares occurred in the treatment group, compared with 16% on placebo. In the earlier study with Abbott, 21% of patients with high-affinity antibodies on placebo experienced flares, compared with 8% on the drug. David Wallace, a rheumatologist at the University of California, Los Angeles, who participated in the trial, attributes the placebo difference to the immunosuppressant mycophenolate, which came on the market between the trials. Approved for patients with organ transplants, doctors quickly began experimenting with it in lupus. Eleven percent of those in the phase III trial were on the drug, he says.

Given the mixed results, LJPC agreed in August with FDA on the design of a larger phase IV postmarketing study, which would include about 600 people and seek to confirm the clinical benefit of the drug. Under a regulation designed to encourage development of drugs for life-threatening illnesses with few remedies, FDA could have approved LJP 394 if the company agreed to conduct the follow-up study. But on 14 October, the agency rejected the company's new drug application, saying another trial was needed.

Since then, LJPC and lupus specialists have met with FDA a half-dozen times, lobbying the agency to reconsider. "What's the downside" of approving LJP 394?, asks Jill Buyon, a lupus specialist at New York University Medical Center who until last fall was a paid consultant for LJPC. She and Engle say the company would pull the drug off the market if it failed in the postmarketing trial.

But Wofsy, who says he doesn't necessarily oppose approval, worries that putting LJP 394 on the market could complicate testing of other therapies for lupus-induced kidney disease. "Would we deny the drug to anyone who wanted it?" he asks. Doctors would be hard pressed to do so, given that no one knows which lupus patients stand to benefit from it.

After a decades-long drought, there are now roughly 10 drugs for lupus in early clinical trials, says Joan Merrill of the Oklahoma Medical Research Foundation in Oklahoma City, who has consulted for LJPC and is the medical director of the Lupus Foundation of America in Washington, D.C. So why such anxiety about abandoning LJP 394? The other drugs might falter, says Merrill, and LJP 394 is possibly "the safest one of all. … We want these things studied," she adds, "until we know they don't work."

—JENNIFER COUZIN

CREDITS (TOP TO BOTTOM): COURTESY OF LUPUS FOUNDATION OF AMERICA INC.; LJPC

SUMO Wrestles Its Way to Prominence in the Cell

The small protein SUMO is turning out to have as many roles in the cell as its better-known cousin, ubiquitin

The protein known as SUMO comes from an illustrious family. Its cousin ubiquitin has long been a star in cell biology: Researchers have shown that it is a key regulator of numerous cellular activities, controlling everything from protein degradation and gene expression to the cell division cycle. Ubiquitin is so renowned, in fact, that its discoverers were awarded last year's Nobel Prize in chemistry (*Science*, 15 October 2004, p. 400). During ubiquitin's ascent, SUMO remained in the shadows. Recently, however, SUMO has begun making a name of its own.

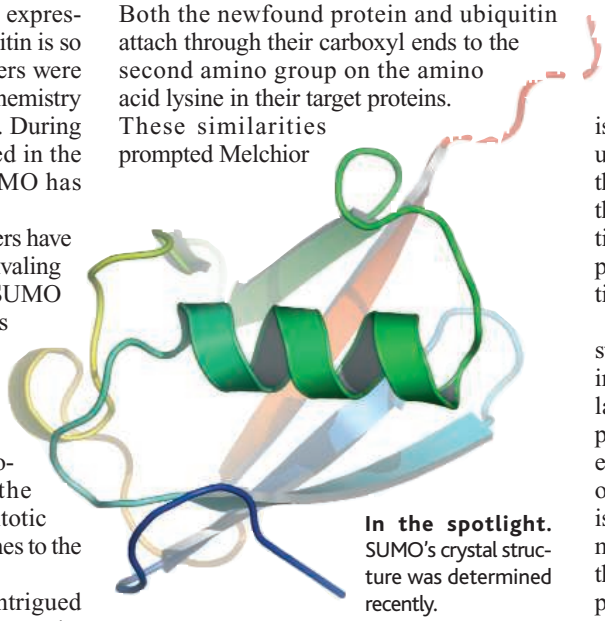
Over the past few years, researchers have implicated it in a range of activities rivaling those of ubiquitin itself. Although SUMO can operate throughout the cell, its actions seem to be concentrated in the nucleus. The molecule has left its fingerprints on many nuclear functions, including gene transcription, DNA repair, the transport of proteins and RNAs into and out of the nucleus, and the building of the mitotic spindle that draws sets of chromosomes to the opposite ends of a dividing cell.

Physicians may one day be as intrigued with SUMO as cell biologists are now. The protein seems to help some viruses infect cells, making it a possible target for antiviral therapies. It may also be involved in neurodegenerative diseases such as Huntington's and Alzheimer's. "There is exciting biology coming out of SUMO research," says Van Wilson of the Texas A&M University System Health Science Center in College Station, who leads one of the groups that have linked SUMO to viral infectivity.

A late start

Although ubiquitin was discovered more than 25 years ago, SUMO eluded detection until 1997, when two groups stumbled on it more or less simultaneously. Both teams, one including Frauke Melchior and Larry Gerace of the Scripps Research Institute in La Jolla, California, and the other including Michael Matunis, who was then working in Gunter Blobel's lab at Rockefeller University in New York City, were studying a protein called RanGAP1 that had been implicated in both nuclear transport and the control of mitosis.

The researchers found that cells contain two forms of RanGAP1, one of which weighs some 20 kilodaltons more than the other. Further analysis showed that the larger form carries an attachment—a 97–amino-acid protein that turned out to resemble ubiquitin in its shape and in the way it connects to RanGAP1. Both the newfound protein and ubiquitin attach through their carboxyl ends to the second amino group on the amino acid lysine in their target proteins. These similarities prompted Melchior and her colleagues to dub the new protein SUMO, which stands for small ubiquitin-like modifier.



In the spotlight. SUMO's crystal structure was determined recently.

Serendipity played a big role in SUMO's discovery. "We got really lucky," says Melchior, who is currently moving her lab from the Max Planck Institute for Biochemistry in Martinsried, Germany, to the University of Göttingen. She explains that RanGAP1 is the only protein in which SUMO stays put when cells are broken apart for analysis. All others rapidly lose their SUMO tags. "I think that may be why [SUMO] was overlooked for so long," she says.

Once SUMO was identified, however, it opened the floodgates. The protein plus two nearly identical relatives found later—dubbed SUMO2 and -3—have since turned up on numerous additional proteins, most of which are located in or around the nucleus. More often than not, researchers encountered a SUMO accidentally, while studying the regulation of some fundamental cell process, such as gene transcription or cell division. "SUMO is popping up in every place you look," says J. Lawrence Marsh of the University of California (UC), Irvine, who is investigating a possible role for the protein in neurodegeneration.

Proteomics studies performed in the last several months have expanded the roster of sumoylated proteins even further. Mark Hochstrasser, whose team at Yale University is one of several performing such analyses, says that the total in yeast, the preferred organism for the work so far, is now up to 150. "I'm sure the number is much higher in mammalian cells," he predicts.

What's it doing?

Although identifying SUMO-adorned proteins is now easy, figuring out exactly what the modification does has proved to be more of a challenge. One thing for sure is that the SUMO tag does not do what ubiquitin addition to proteins often does: mark them for destruction by a cell structure called the proteasome. In fact, there are a few situations in which SUMO modification protects proteins from degradation by blocking addition of a ubiquitin tag.

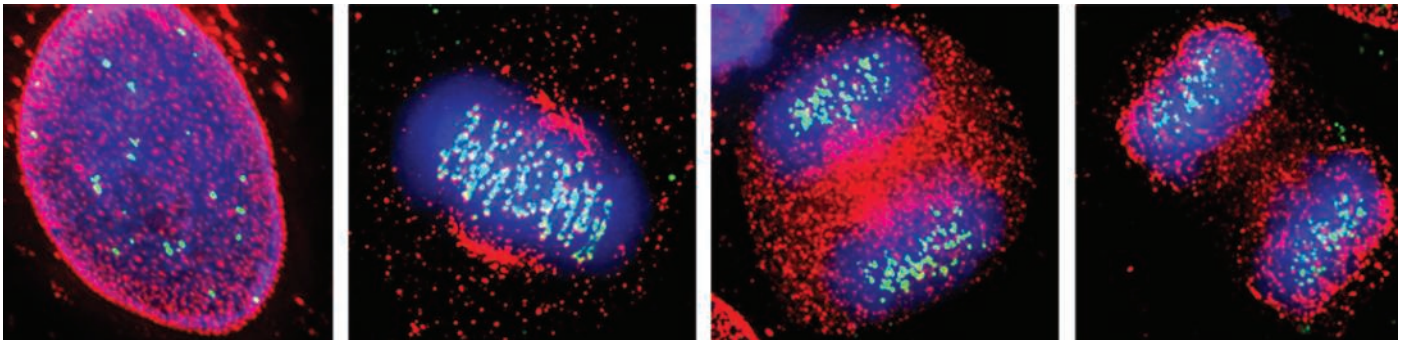
Researchers have been building a circumstantial case that SUMO is involved in directing protein movements in the cell, particularly the transport of proteins through the pores of the nuclear envelope. That idea emerged early with the discovery of SUMO1 on RanGAP1. Both the Matunis team, which is now at Johns Hopkins University in Baltimore, Maryland, and that of Melchior showed that unmodified RanGAP1, which is located primarily in the cell cytoplasm, moves when sumoylated to the tiny fibrils that project from the outer side of nuclear pores.

Since then, components of the machinery that sumoylates proteins have also turned up at the pore. For example, Matunis and his colleagues have located a protein called Ubc9 at the pore. In the first step of the sumoylation reaction, which is similar to how ubiquitin is added to proteins, SUMO forms a high-energy bond to the so-called E1 activating protein. Then SUMO is transferred to an E2 conjugating protein—Ubc9—and from there it's joined to its ultimate protein target with the aid of an E3 ligase, which is needed for target recognition.

Researchers were somewhat surprised to learn that the machinery includes an E3 ligase because in test tube experiments E1 and E2 seemed sufficient to pin SUMO on proteins. But in the past 4 years, researchers in several labs have identified a half-dozen or so E3s that do this job. As shown by Melchior's team, working with Anne Dejean and her colleagues at the Pasteur Institute in Paris, these include a protein called RanBP2/Nup358, which is located at the nuclear pore and is known to bind RanGAP1. The supposition is

CREDIT: L. M. LOIS AND C. D. LIMA, *EMBO JOURNAL*, 20 JANUARY 2005

Published by AAAS



As the cell divides. Before mitosis begins (left panel), the sumoylated protein RanGAP1 (red) is located mostly at the nuclear pores. But as mitosis gets under way, it moves to the kinetochores and mitotic spindle (middle panels) and then redistributes to the pores of the new nuclei after the daughter cells separate.

that RanBP2 is involved in sumoylating RanGAP1 and other proteins at the pore, although that has not been proven.

In addition, Hochstrasser and others have identified protease enzymes that can remove SUMO from proteins. "These are reversible modifications," Hochstrasser says. The situation, which parallels that for addition of other protein-regulating modifiers such as ubiquitin and phosphate, provides for dynamic control by the cell of the modified proteins. At least one of these SUMO-stripping proteases has also been located at nuclear pores by the Matunis team and by Mary Dasso and her colleagues at the National Institute of Child Health and Human Development (NICHD) in Bethesda, Maryland.

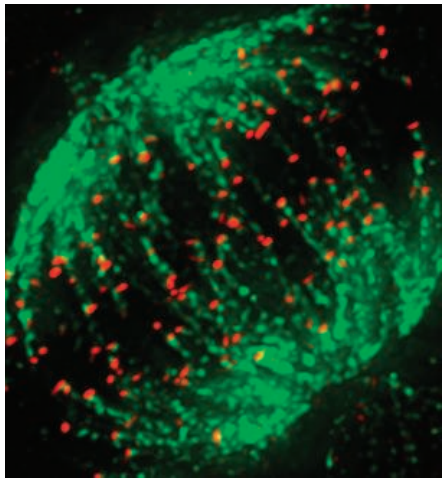
The presence at such pores of the various enzymes involved in SUMO addition and removal raises the possibility that sumoylation serves as a kind of gatekeeper, regulating traffic into and out of the nucleus. This may be the case for a nuclear enzyme called histone deacetylase 4 (HDAC4), which removes acetyl groups from histone proteins in the chromatin. This action allows the DNA to condense and thus has the effect of repressing gene transcription.

Three years ago, Dejean, Melchior, and their colleagues filled in some details suggesting how HDAC4 might work. They showed that it must be sumoylated to produce its full gene-suppressing activity and that the nuclear pore protein, RanBP2, promotes HDAC4 sumoylation. That suggests that HDAC4 picks up its SUMO tag as it moves into the nucleus. There's still room for uncertainty, however. Sumoylating enzymes are present both inside the nucleus and in cytoplasm, leaving open the possibility that HDAC4 picks up its SUMO tag elsewhere.

Wherever sumoylation takes place, it can have important functional consequences, particularly in regulating gene activity. Perhaps 50% of the proteins altered by the tag are transcription factors that are involved in turning genes on or off.

In most cases, adding SUMO results in lowered activity of the target genes. Researchers have shown this by, for example, mutating the sumoylation site on the transcription factors, thus preventing SUMO attachment. The result: increased gene expression. Again, though, things get somewhat murky when it comes to the mechanism by which this inhibition happens. "The problem is that SUMO can regulate so many functions of proteins," says Kevin Sarge of the University of Kentucky in Lexington.

Indeed, the protein's role in transcription is complex. SUMO-driven inhibition of gene expression may occur in several different ways—more than one of which may be operating at a time. Most transcription factors work with several protein partners, and sumoylation may interfere with their interactions. Or, as has been shown for several nuclear proteins including some transcription factors, sumoylation can direct proteins to so-called PML nuclear bodies, small particles located in the nucleus. This may take them out of action, perhaps simply by sequestering them away from the DNA.



Metaphase close-up. As shown by the red staining, sumoylated RanGAP1 localizes to the kinetochores, which attach the chromosomes to the mitotic spindle.

Although sumoylation of transcription factors usually results in decreased gene expression, occasionally the opposite occurs. Sarge and his colleagues provide some intriguing examples. They have been studying the activities of heat shock factors (HSFs), proteins that protect the cell against heat and other stresses by turning up the activity of a variety of protective genes.

In work done a few years ago with Matunis and his colleagues, the Sarge team showed that HSF1 is sumoylated in response to stress and that this leads to activation of HSF1's target genes. In this case, adding SUMO may increase HSF1's binding to DNA.

Something similar happens with HSF2, although here the sumoylation trigger is not stress but the cell cycle transition from the second growth phase to actual cell division. When cells are preparing to divide, they compact most of the DNA of their chromosomes with the aid of an enzyme called condensin. In order for cells to function, however, essential genes have to be kept open, and the new work indicates that SUMO plays a role in this "bookmarking" of critical points in DNA. The Kentucky team reported in the 21 January issue of *Science* (p. 421) that sumoylated HSF2 binds both to a target gene and to CAPG, a subunit of condensin, and then draws in an enzyme that inactivates the condensing enzyme. As a result, the DNA stays open in the gene's vicinity.

Further experiments showed the importance of HSF2 sumoylation for cell survival. When Sarge and his colleagues blocked the synthesis of HSF2 with an inhibitory RNA, they found that control cells could withstand an elevated temperature of 43°C, but that many of the cells carrying the inhibitory RNA died at that temperature.

Protecting the genome

SUMO's roles in the nucleus go far beyond regulating protein transport and gene transcription. The protein is also needed for normal separation of the chromosomes during mitosis and is involved in repairing damaged DNA. Researchers have found that mutations in SUMO genes themselves or in genes for enzymes involved in adding or removing the protein from its targets lead to abnormal cell division and increased susceptibility to DNA-damaging agents.

CREDITS (TOP TO BOTTOM): J. JOSEPH *ET AL.*, *THE JOURNAL OF CELL BIOLOGY* **156**, 595 (2002); J. JOSEPH, S.-H. TAN, T. KARPOVA, J. MCNALLY, AND M. DASSO/NIH

Getting it on. SUMO addition occurs in three steps. In the first, the protein is attached to the E1 activating enzyme with the aid of energy released from ATP hydrolysis. SUMO is then transferred to the E2 conjugating protein (Ubc9), and from there an E3 ligase directs it to its target protein.

During cell division, the duplicated daughter chromosomes are joined together at their central regions, the centromeres, before they ultimately separate. The evidence so far indicates that sumoylation may help signal the separation. Working with yeast, Stephen Elledge of Baylor College of Medicine in Houston, Texas, Nancy Kleckner of Harvard University in Cambridge, Massachusetts, and their colleagues discovered that mutating the gene for one of the SUMO-removing proteases results in premature chromosome separation. The researchers have evidence that this effect involves topoisomerase II (Top2), an enzyme known to regulate chromosome structure during mitosis.

According to the model they proposed, SUMO is constantly being added to and removed from Top2 by the appropriate enzymes. The unsumoylated version is the one that helps maintain chromosome cohesion, perhaps through its effects on chromosome structure. But when tagged with SUMO, Top2 can no longer sustain the cohesion, allowing chromosome separation. Thus, when a mutation inactivates the protease that should remove SUMO from Top2, the sumoylated version will accumulate at the expense of the unsumoylated form. The result: Chromosomes separate prematurely.

Consistent with this model, Dasso and her NICHD colleagues have recently found that Top2 is heavily sumoylated during mitosis in frog eggs. And as expected, preventing that sumoylation blocked chromosome separation.

Sumoylation might also be involved in another critical event involving the centromere: formation of the kinetochore that attaches the chromosomes to the microtubule fibers of the mitotic spindle that draw the separating chromosomes to the opposite ends of the cell. Researchers have found that SUMO modifies several kinetochore and centromere proteins. And Dasso's team has found that in cultured human cells, addition of SUMO to RanGAP1 is what targets the protein to the kinetochore and mitotic spindle. The presence of RanGAP1, which activates one of the enzymes involved in spindle assembly, "may be needed for kinetochore integrity and microtubule attachment," Dasso suggests.

Further evidence for that idea comes from Brian Burke's team at the University of Florida, Gainesville. These researchers found that depletion of RanBP2, the SUMO E3 ligase, results in abnormalities in kinetochore structure and thus in mitosis. This might be because RanBP2 binds RanGAP1 at the kinetochore just as it does at the nuclear pore, or because it is needed for sumoylation of RanGAP1, or both.

SUMO may also pitch in to help regulate DNA repair. Because DNA is constantly subject to damage, either through errors in replication or by exposure to chemicals or radiation, a cell needs to maintain an effective repair machinery. Researchers have found that sumoylation regulates the activities of several proteins involved in DNA repair. These include p53, sometimes called the "guardian of the genome" because of the key role it plays in DNA repair, and a protein called PCNA (proliferating cell nuclear antigen).

Stefan Jentsch and his colleagues at the Max Planck Institute for Biochemistry showed that ubiquitin addition to PCNA promotes its DNA-repairing activity. In contrast, sumoylation inhibits that activity, apparently because it goes on at the same site, thereby precluding ubiquitin addition. The Martinsreid workers speculate that SUMO may direct PCNA to another function, perhaps in DNA replication.

Disease links

With sumoylation apparently involved in so many cellular activities, it's not surprising that it may be relevant to diseases as well. There are hints that it is involved in the pathology of neurodegenerative diseases, including Huntington's, which is caused by mutations in the Huntingtin protein (Htt). Last spring, UC Irvine's Marsh and his colleagues reported that sumoylation of Htt increases its neurotoxicity (*Science*, 2 April 2004, p. 100) both in cultured human neurons and in a fruit fly model of Huntington's disease.

Even certain viruses, such as human papillomavirus and the herpesviruses, may utilize a cell's SUMO for their own nefarious purposes. In some cases, sumoylation of viral proteins targets them to the nucleus so that they can take over the cell's replication machinery, thus allowing viral reproduction. For example, Wilson and his colleagues found that blocking sumoylation of a human papillomavirus protein causes it to lose the ability to activate viral replication.

Viruses may also aid their cause by interfering with the cell's sumoylation machinery. A team including Julio Draetta and Susanna Chiocca of the European Institute of Oncology in Milan and Ronald Hay of the University of St. Andrews in the United Kingdom reported in the November issue of *Molecular Cell* that Gam1, a protein from an avian adenovirus, inactivates the SUMO E1 protein in cultured human cells, thus totally blocking sumoylation.

Because SUMO addition to transcription factors tends to inhibit transcription, the result is an overall increase in gene expression, presumably including those of the virus. "Viruses are going to affect host sumoylation with the goal of making an environment in the cell that is favorable for viral replication," Wilson says.

Although much remains to be learned about SUMO and its actions, it's already clear that its discovery has opened numerous lines of investigation. Researchers are learning that even a small protein is able to throw its weight around in the cell.

–JEAN MARX

Calorie Count Reveals Neandertals Out-Ate Hardiest Modern Hunters

NEW YORK CITY—Top Neandertal experts gathered from 27 to 29 January for an invitation-only conference sponsored by New York University and the Max Planck Institute for Evolutionary Anthropology in Leipzig.

Cartoons, B-movies, and anthropologists agree that Neandertals were a husky tribe. But how much fuel was needed to power those stocky, powerful frames? Scientists have speculated that supporting such massive bodies in the chill of glacial Europe required a hefty dose of calories and perhaps oxygen to burn them; the need for increased oxygen, in turn, might have spurred the evolution of Neandertals' large chests, which presumably enclosed capacious lungs. At the meeting, paleoanthropologist Steve Churchill of Duke University in Durham, North Carolina, unveiled numbers to test those ideas.

Action man. Half-size model helped compute Neandertals' turbocharged metabolism.

For living humans, physiologists have developed equations that relate parameters such as size, skin surface area, and basal metabolic rate (BMR, or the number of calories burned to maintain body temperature at rest). To tailor the equations to short-limbed, big-muscled Neandertals, Churchill created a half-size Neandertal model, proportioned after a famous skeleton from La Ferrassie in France. He plastered the model's surface with a silicone rubber peel, digitized the peel, scaled up his results, and concluded that an 84-kilogram, 171-cm-tall male Neandertal was wrapped in 2.1 square meters of skin.

Neandertal data in hand, Churchill used equations from modern human physiology to estimate a male Neandertal BMR of about 2000 calories per day, about 25% more than the average for a modern American male. Then he assumed that Neandertals were about as active as modern people who hunt game near an ice front, namely living Inuit hunters, whose activity consumes about 2 to 2.5 times the calories spent in basal metabolism. Churchill concluded that a male Neandertal needed 4500 to 5040 kilocalories per day to survive; for comparison, Inuit people consume about 3000 to 4000.

Chemical isotopes in their bones indicate that Neandertals were the original, extreme Atkins dieters, dining almost exclusively on meat. That means that a male Neandertal would have needed to nosh one healthy caribou per month. "That's two kilos of caribou a day," says Churchill. "That's a lot of meat." A mixed band of 10 Neandertals might have needed two caribou a week, he said.

Supersized servings might also have led to beefy oxygen requirements and could help explain Neandertals' large chests, Churchill says. Using equations relating energy expenditure to oxygen intake, he concludes that Neandertals at rest respired at an average rate of 19 liters of air per minute—two or three times as much as modern people breathe at rest. So part of Neandertals' generous lung capacity may have gone simply to power resting respiration, says Churchill. Bursts of activity such as hunting might have required even more breathing power, he says.

Researchers of diverse backgrounds praised Churchill's work. "That pop-up Neandertal is very direct and absolutely the right way to do it," says paleoanthropologist Milford Wolpoff of the University of Michigan, Ann Arbor. "He's not saying Neandertals are Eskimos, but his estimates are compatible with real data from real people. To me that's exciting."

Churchill adds that his calorie calculations show that at times Neandertals may have come perilously close to the edge of survival, especially at the end of winter when food was scarce. "Their caloric budgets must have been tight," he says. He notes that many Inuit undergo yearly fasts and that Neandertal teeth are "full of defects that indicate periodic starvation." Periods of winter stress fit with other data on the Neandertals as well as studies of modern hunter-gatherers, agrees archaeologist Alison Brooks of George Washington University in Washington, D.C. "The hunter-gatherer life in the past was very precarious. Even modern hunter-gatherers often lose 10% of their weight in the bad season, and that can stop ovulation and reproduction." The hunting-dependent Neandertals would have "had the fewest calories available at the coldest time of year; it must have been very stressful," she says.

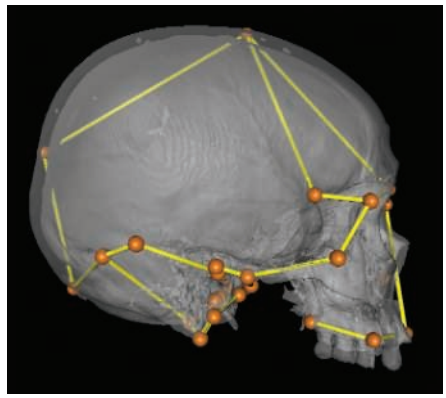
All the same, Neandertals apparently thrived for 600,000 years in Europe's harsh glacial climes, Wolpoff points out: "You can't have a population close to the edge for that long."

"Their overall adaptive strategy was successful," Churchill agrees. "But it was an energetically expensive adaptation."

Faces May Lie When Skulls Tell Tales

For decades anthropologists comparing fossils have argued bitterly about whether similarities are due to family resemblance or to convergent evolution. For example, both living Europeans and Neandertals have high-bridged, projecting noses, and a few researchers have cited this as evidence of Neandertal ancestry. But others say big schnozzes may merely reflect independent adaptations to Europe's chilly weather.

At the meeting, Katerina Harvati and Tim Weaver of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, presented a new way to sort out how genetics and environment affect three parts of the human skull: the face, the braincase or vault, and the temporal bone,



Power points. Some parts of the skull give more evolutionary information than others.

CREDITS (TOP TO BOTTOM): S. CHURCHILL/DUKE UNIVERSITY; KATERINA HARVATI/MAX PLANCK INSTITUTE FOR EVOLUTIONARY ANTHROPOLOGY

Published by AAAS

Calorie Count Reveals Neandertals Out-Ate Hardiest Modern Hunters

NEW YORK CITY—Top Neandertal experts gathered from 27 to 29 January for an invitation-only conference sponsored by New York University and the Max Planck Institute for Evolutionary Anthropology in Leipzig.

Cartoons, B-movies, and anthropologists agree that Neandertals were a husky tribe. But how much fuel was needed to power those stocky, powerful frames? Scientists have speculated that supporting such massive bodies in the chill of glacial Europe required a hefty dose of calories and perhaps oxygen to burn them; the need for increased oxygen, in turn, might have spurred the evolution of Neandertals' large chests, which presumably enclosed capacious lungs. At the meeting, paleoanthropologist Steve Churchill of Duke University in Durham, North Carolina, unveiled numbers to test those ideas.

Action man. Half-size model helped compute Neandertals' turbocharged metabolism.

For living humans, physiologists have developed equations that relate parameters such as size, skin surface area, and basal metabolic rate (BMR, or the number of calories burned to maintain body temperature at rest). To tailor the equations to short-limbed, big-muscled Neandertals, Churchill created a half-size Neandertal model, proportioned after a famous skeleton from La Ferrassie in France. He plastered the model's surface with a silicone rubber peel, digitized the peel, scaled up his results, and concluded that an 84-kilogram, 171-cm-tall male Neandertal was wrapped in 2.1 square meters of skin.

Neandertal data in hand, Churchill used equations from modern human physiology to estimate a male Neandertal BMR of about 2000 calories per day, about 25% more than the average for a modern American male. Then he assumed that Neandertals were about as active as modern people who hunt game near an ice front, namely living Inuit hunters, whose activity consumes about 2 to 2.5 times the calories spent in basal metabolism. Churchill concluded that a male Neandertal needed 4500 to 5040 kilocalories per day to survive; for comparison, Inuit people consume about 3000 to 4000.

Chemical isotopes in their bones indicate that Neandertals were the original, extreme Atkins dieters, dining almost exclusively on meat. That means that a male Neandertal would have needed to nosh one healthy caribou per month. "That's two kilos of caribou a day," says Churchill. "That's a lot of meat." A mixed band of 10 Neandertals might have needed two caribou a week, he said.

Supersized servings might also have led to beefy oxygen requirements and could help explain Neandertals' large chests, Churchill says. Using equations relating energy expenditure to oxygen intake, he concludes that Neandertals at rest respired at an average rate of 19 liters of air per minute—two or three times as much as modern people breathe at rest. So part of Neandertals' generous lung capacity may have gone simply to power resting respiration, says Churchill. Bursts of activity such as hunting might have required even more breathing power, he says.

Researchers of diverse backgrounds praised Churchill's work. "That pop-up Neandertal is very direct and absolutely the right way to do it," says paleoanthropologist Milford Wolpoff of the University of Michigan, Ann Arbor. "He's not saying Neandertals are Eskimos, but his estimates are compatible with real data from real people. To me that's exciting."

Churchill adds that his calorie calculations show that at times Neandertals may have come perilously close to the edge of survival, especially at the end of winter when food was scarce. "Their caloric budgets must have been tight," he says. He notes that many Inuit undergo yearly fasts and that Neandertal teeth are "full of defects that indicate periodic starvation." Periods of winter stress fit with other data on the Neandertals as well as studies of modern hunter-gatherers, agrees archaeologist Alison Brooks of George Washington University in Washington, D.C. "The hunter-gatherer life in the past was very precarious. Even modern hunter-gatherers often lose 10% of their weight in the bad season, and that can stop ovulation and reproduction." The hunting-dependent Neandertals would have "had the fewest calories available at the coldest time of year; it must have been very stressful," she says.

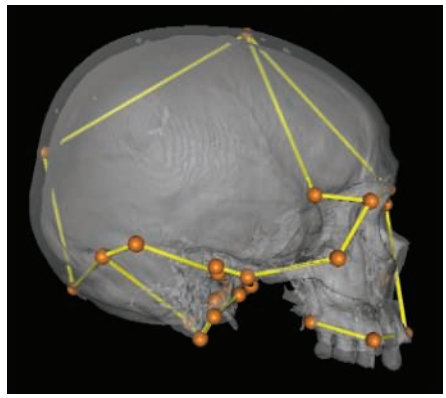
All the same, Neandertals apparently thrived for 600,000 years in Europe's harsh glacial climes, Wolpoff points out: "You can't have a population close to the edge for that long."

"Their overall adaptive strategy was successful," Churchill agrees. "But it was an energetically expensive adaptation."

Faces May Lie When Skulls Tell Tales

For decades anthropologists comparing fossils have argued bitterly about whether similarities are due to family resemblance or to convergent evolution. For example, both living Europeans and Neandertals have high-bridged, projecting noses, and a few researchers have cited this as evidence of Neandertal ancestry. But others say big schnozzes may merely reflect independent adaptations to Europe's chilly weather.

At the meeting, Katerina Harvati and Tim Weaver of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, presented a new way to sort out how genetics and environment affect three parts of the human skull: the face, the braincase or vault, and the temporal bone,



Power points. Some parts of the skull give more evolutionary information than others.

CREDITS (TOP TO BOTTOM): S. CHURCHILL/DUKE UNIVERSITY; KATERINA HARVATI/MAX PLANCK INSTITUTE FOR EVOLUTIONARY ANTHROPOLOGY

Published by AAAS

which comprises the temple, the ear, and the upper jaw joint. "People have said, 'This or that feature is best to track population history,' but it's never really been tested," said Harvati. With samples from individuals in 10 populations throughout the world, Harvati and Weaver compared three kinds of data: differences in skull morphology, or shape; genetic differences taken from Stanford University geneticist Luigi Cavalli-Sforza's published global database; and climatic differences, as represented by latitude and mean temperature.

They found that morphological differences did indeed correlate with genetic ones in each part of the skull. But the shape of the face was also associated with climate. For example, Greenlanders and northern Europeans, although relatively distant genetically, both tend to have flat faces.

In contrast, the vault did not reflect climate but tracked genes closely. For example, Syrians, Italians, and Greeks "clustered together beautifully," both genetically and in vault shape, revealing recent population history, Harvati says. The temporal bone tracked more ancient population history, she says. Only in this part of the skull were Africans distinct from all other populations, mapping the most ancient split seen in the genetic data. "So if you're looking deep into time, you probably want to use the temporal bone and avoid the face, because the face reflects a complex mix of genes and climate," Harvati says. Their analysis of temporal bone shape shows that living and Upper Paleolithic modern humans cluster together but that Neandertals are quite distinct from both, suggesting that they are indeed different species.

Many at the meeting praised what paleoanthropologist Steve Churchill of Duke University in Durham, North Carolina, called Harvati and Weaver's "right-headed approach." "I'm full of admiration for [Harvati's] work," said paleoanthropologist Chris Stringer of the Natural History Museum in London. Several researchers pointed out ways to improve the analysis, however, suggesting everything from better genetic data sets to more precise climatic data. And they noted that many anthropologists already rely on the temporal bone—and steer clear of the face—when sorting out evolutionary relationships. All the same, says paleoanthropologist Ian Tattersall of the American Museum of Natural History in New York City, "this is a very imaginative approach, and it's a harbinger for future advances."

–ELIZABETH CULOTTA

CREDIT: KEN MOWBRAY/AMERICAN MUSEUM OF NATURAL HISTORY

The Question of Sex

For 150 years, members of *Homo sapiens* have gazed on the bones of Neandertals and wondered, "Was this one of us?" At a recent meeting in New York, many paleoanthropologists—although not all—answered "No." Yet even partisans committed to the notion that Neandertals and moderns were separate species agreed that when the groups met, at least a bit of what Ian Tattersall of the American Museum of Natural History in New York City calls "Pleistocene hanky-panky" probably took place.

Gene swappers? Most researchers say modern humans (right) and Neandertals got together—but not often.

Carbon-14 dating of fossils and artifacts suggests that Neandertals and modern humans coexisted for several thousand years in Western Europe, after moderns swept in from Africa and before Neandertals vanished about 28,000 years ago. The minority of researchers who think Neandertals and moderns belonged to a single species have no doubt about what happened next: "What do we do when we encounter anyone? We trade mates and culture," says Milford Wolpoff of the University of Michigan, Ann Arbor, who has long argued for a single interbreeding human population. "The archaeological record is clearly showing us that these groups are trading ideas, which almost certainly means they're trading mates."

Indeed, the idea of thousands of years of chaste coexistence is too much of a stretch even for many experts who believe Neandertals were a separate species. "If you're counting on humans *not* to mate, you'll be very disappointed," warned paleoanthropologist Trent Holliday of Tulane University in New Orleans, Louisiana. In his presentation, Holliday argued that any attempted gene-swapping could well have succeeded. By his count, about 1/3 of known mammalian hybrids are fertile. They include crosses between mule deer and white-tailed deer, lynx and bobcat, and many others. Primatologist Cliff Jolly of New York University, speaking from the audience, added a crucial primate example: olive and hamadryas baboons of Ethiopia, visibly distinct forms with different social structures. According to mitochondrial DNA (mtDNA), their ancestors diverged about 300,000 to 500,000 years ago, roughly the same time modern humans and Neandertals evolved in separate lineages. In the wild, the baboons freely interbreed within a narrow hybrid zone. "With them as a primate parallel, you'd expect that Neandertals and moderns would have been reproductively compatible," says Jolly.

Yet the ancient mtDNA so far gathered from a handful of Neandertals is distinct from that of both early and living modern humans. Jolly and others suggest that behavioral or cultural differences might have kept the gene pools of modern humans and Neandertals mostly distinct even in the face of some mating. Even so, they say, that doesn't mean abstinence worked perfectly. "Neandertals and moderns can be regarded as distinct species, but that does not mean that they were completely reproductively isolated," says Chris Stringer of the Natural History Museum in London, a longtime advocate of the notion that modern humans replaced the Neandertal species. "The point that came out [at the meeting] is that you can have both: distinct species, and some reproduction."

The real question, said paleoanthropologist Jean-Jacques Hublin of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, is whether that reproduction affected later populations of *Homo sapiens*. "As for sex in the past: They [Neandertals and moderns] did it. I believe that. But does this have any biological relevance? No." Hublin, Stringer, and others at the conference see no evidence, from fossils or ancient DNA, that Neandertals are part of modern humans' ancestry. Thus they argue that hybridization must have been quite limited.

Jolly notes that a few genes that were highly adaptive for the local environment might have found their way from Neandertals to modern humans. Genes for pale skin color, for example, are advantageous in the sun-starved north. Wolpoff and a few others go further. They emphasize that even the geneticists admit that current mtDNA data cannot completely rule out a Neandertal contribution, and they cited a few Upper Paleolithic fossils that may show signs of Neandertal traits. Paleolithic hybrids may have bridged two species, but the question of their impact remains as divisive as ever. **–E.C.**

Published by AAAS

which comprises the temple, the ear, and the upper jaw joint. "People have said, 'This or that feature is best to track population history,' but it's never really been tested," said Harvati. With samples from individuals in 10 populations throughout the world, Harvati and Weaver compared three kinds of data: differences in skull morphology, or shape; genetic differences taken from Stanford University geneticist Luigi Cavalli-Sforza's published global database; and climatic differences, as represented by latitude and mean temperature.

They found that morphological differences did indeed correlate with genetic ones in each part of the skull. But the shape of the face was also associated with climate. For example, Greenlanders and northern Europeans, although relatively distant genetically, both tend to have flat faces.

In contrast, the vault did not reflect climate but tracked genes closely. For example, Syrians, Italians, and Greeks "clustered together beautifully," both genetically and in vault shape, revealing recent population history, Harvati says. The temporal bone tracked more ancient population history, she says. Only in this part of the skull were Africans distinct from all other populations, mapping the most ancient split seen in the genetic data. "So if you're looking deep into time, you probably want to use the temporal bone and avoid the face, because the face reflects a complex mix of genes and climate," Harvati says. Their analysis of temporal bone shape shows that living and Upper Paleolithic modern humans cluster together but that Neandertals are quite distinct from both, suggesting that they are indeed different species.

Many at the meeting praised what paleoanthropologist Steve Churchill of Duke University in Durham, North Carolina, called Harvati and Weaver's "right-headed approach." "I'm full of admiration for [Harvati's] work," said paleoanthropologist Chris Stringer of the Natural History Museum in London. Several researchers pointed out ways to improve the analysis, however, suggesting everything from better genetic data sets to more precise climatic data. And they noted that many anthropologists already rely on the temporal bone—and steer clear of the face—when sorting out evolutionary relationships. All the same, says paleoanthropologist Ian Tattersall of the American Museum of Natural History in New York City, "this is a very imaginative approach, and it's a harbinger for future advances."

–ELIZABETH CULOTTA

CREDIT: KEN MOWBRAY/AMERICAN MUSEUM OF NATURAL HISTORY

The Question of Sex

For 150 years, members of *Homo sapiens* have gazed on the bones of Neandertals and wondered, "Was this one of us?" At a recent meeting in New York, many paleoanthropologists—although not all—answered "No." Yet even partisans committed to the notion that Neandertals and moderns were separate species agreed that when the groups met, at least a bit of what Ian Tattersall of the American Museum of Natural History in New York City calls "Pleistocene hanky-panky" probably took place.

Gene swappers? Most researchers say modern humans (right) and Neandertals got together—but not often.

Carbon-14 dating of fossils and artifacts suggests that Neandertals and modern humans coexisted for several thousand years in Western Europe, after moderns swept in from Africa and before Neandertals vanished about 28,000 years ago. The minority of researchers who think Neandertals and moderns belonged to a single species have no doubt about what happened next: "What do we do when we encounter anyone? We trade mates and culture," says Milford Wolpoff of the University of Michigan, Ann Arbor, who has long argued for a single interbreeding human population. "The archaeological record is clearly showing us that these groups are trading ideas, which almost certainly means they're trading mates."

Indeed, the idea of thousands of years of chaste coexistence is too much of a stretch even for many experts who believe Neandertals were a separate species. "If you're counting on humans *not* to mate, you'll be very disappointed," warned paleoanthropologist Trent Holliday of Tulane University in New Orleans, Louisiana. In his presentation, Holliday argued that any attempted gene-swapping could well have succeeded. By his count, about 1/3 of known mammalian hybrids are fertile. They include crosses between mule deer and white-tailed deer, lynx and bobcat, and many others. Primatologist Cliff Jolly of New York University, speaking from the audience, added a crucial primate example: olive and hamadryas baboons of Ethiopia, visibly distinct forms with different social structures. According to mitochondrial DNA (mtDNA), their ancestors diverged about 300,000 to 500,000 years ago, roughly the same time modern humans and Neandertals evolved in separate lineages. In the wild, the baboons freely interbreed within a narrow hybrid zone. "With them as a primate parallel, you'd expect that Neandertals and moderns would have been reproductively compatible," says Jolly.

Yet the ancient mtDNA so far gathered from a handful of Neandertals is distinct from that of both early and living modern humans. Jolly and others suggest that behavioral or cultural differences might have kept the gene pools of modern humans and Neandertals mostly distinct even in the face of some mating. Even so, they say, that doesn't mean abstinence worked perfectly. "Neandertals and moderns can be regarded as distinct species, but that does not mean that they were completely reproductively isolated," says Chris Stringer of the Natural History Museum in London, a longtime advocate of the notion that modern humans replaced the Neandertal species. "The point that came out [at the meeting] is that you can have both: distinct species, and some reproduction."

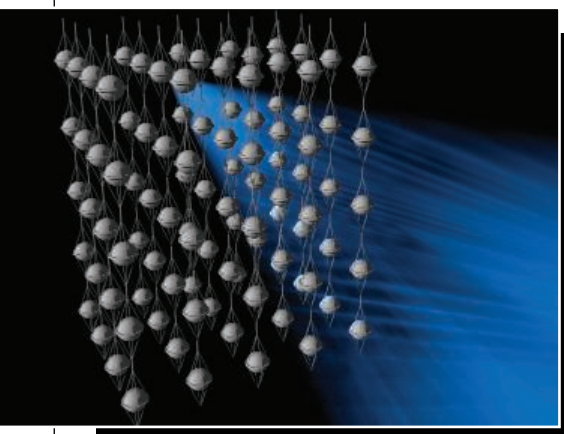
The real question, said paleoanthropologist Jean-Jacques Hublin of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, is whether that reproduction affected later populations of *Homo sapiens*. "As for sex in the past: They [Neandertals and moderns] did it. I believe that. But does this have any biological relevance? No." Hublin, Stringer, and others at the conference see no evidence, from fossils or ancient DNA, that Neandertals are part of modern humans' ancestry. Thus they argue that hybridization must have been quite limited.

Jolly notes that a few genes that were highly adaptive for the local environment might have found their way from Neandertals to modern humans. Genes for pale skin color, for example, are advantageous in the sun-starved north. Wolpoff and a few others go further. They emphasize that even the geneticists admit that current mtDNA data cannot completely rule out a Neandertal contribution, and they cited a few Upper Paleolithic fossils that may show signs of Neandertal traits. Paleolithic hybrids may have bridged two species, but the question of their impact remains as divisive as ever. **–E.C.**

Published by AAAS

Random Samples

Edited by Constance Holden



Neutrinos on the Rocks

Scientists are turning a giant chunk of Antarctic ice into a new window on the universe. Last month they drilled the first hole for IceCube, a detector designed to look for telltale flashes of light generated when ice molecules are hit by high-energy neutrinos from outer space.

Over the next 5 years, engineers plan to use hot water to drill 80 2.4-km-deep holes and drop in long cables, each equipped with 64 light sensors, before the water freezes. Scientists say polar ice—pure, dark, and transparent—is perfect for detecting neutrinos, which are the only particles able to traverse Earth relatively unhindered.

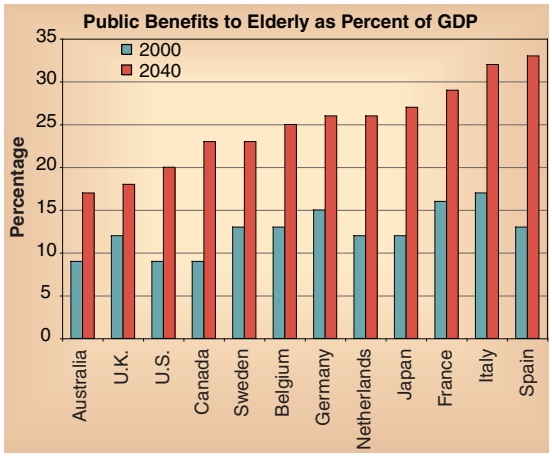
The University of Wisconsin, Madison, is coordinating the IceCube project. Team member Nick van Eijndhoven of Utrecht University in the Netherlands expects the detector to shed new light on the energetic events in distant active galaxies and on mysterious stellar explosions called gamma ray bursts. IceCube will incorporate a small array called AMANDA. "This monster will really do a good job," says van Eijndhoven. "It will dwarf AMANDA within 3 years," eventually covering 60 times the volume of ice.

Wake-Up Call for an Aging Europe

The countries of continental Europe—particularly France, Spain, and Italy—may not be able to support their elderly populations by 2040. So predicts Richard Jackson of the Center for Strategic and International Studies in Washington, D.C., who has computed an "aging vulnerability index" for 12 developed countries based on the growth in the elderly populations and the government benefits they're receiving.

"Everybody expected when I embarked on this project to see Japan" emerge as the worst case, Jackson said during a presentation last month. But in Japan 50% of old people live with their children and their government benefits are "relatively stingy," said Jackson.

Continental Europe is another story. Jackson noted that most Europeans are retired by the time they hit 60—which is only 10 years above population median age in 2050. France has a particularly crushing burden. There, public benefits account for 70% of retirees' incomes—compared with less than 40% in the United States and the United Kingdom. And thanks to policies encouraging early retirement, only 2% of those over 65 are still in the French workforce. Italy and Spain are also heading for trouble, with more retirees than workers by midcentury, forecasts Jackson. He said that even a heavy influx of younger immigrants would not change the picture—"to make a substantial long-term difference, Europe would need a permanent increase of 20 million people a year." But, noted Jackson, there is a way to cut the Gordian knot: "Redefine elderly at 75 and the problem goes away."



Men and Muscle

Are Asian men more comfortable with their bodies than Caucasians? Such questions are of interest to researchers these days with the growth—among men of the West—of so-called muscle dysmorphia, a condition in which people never think they're big or muscular enough.

Several years ago psychiatrist Harrison Pope of McLean Hospital in Belmont, Massachusetts, and colleagues published a paper showing that among a group of U.S. and European men, the "ideal" body type sported some 13 kilograms more muscle than the men had. The men also estimated—erroneously—that women would find that more attractive.

Not the Eastern ideal.

Pope and his student Jeffrey Yang decided to find out if this disconnect was a cultural universal. So they asked 55 young, heterosexual males in Taiwan to choose images that best represented their own bodies and their ideal shapes. These men chose an ideal body with only 2 kilograms more muscle, the authors report in the February *American Journal of Psychiatry*.

The authors also report that a survey of ads in U.S. and Taiwanese women's magazines supported the notion that in Asia, muscles don't make the man. "Asian men were almost never shown undressed," while 43% of Western men were baring pecs or abs.

The results fit with the rarity in Asia of "body-image disorders" and the lack of interest in muscle-building drugs, says Pope. He and Yang speculate that Chinese culture plays a role because it has a subtle ideal of masculinity that emphasizes brains over sheer brawn. Psychologist Jack Darkes of the University of South Florida, Tampa, calls the research "timely" and says it would be interesting to see the results of such a study in Japan, where Westernization is more pervasive.

CREDITS (TOP TO BOTTOM): UW MADISON/ICE CUBE PROJECT; C. J. PHAN/WWW.ASIAFITNESSCOACH.COM; SOURCE: R. JACKSON

Published by AAAS

PEOPLE

Edited by Yudhijit Bhattacharjee

JOBS

New at NSF. Linguist David Lightfoot has been appointed head of the social and behavioral sciences directorate at the National Science Foundation (NSF). The 59-year-old Lightfoot, dean of the graduate school at Georgetown University in

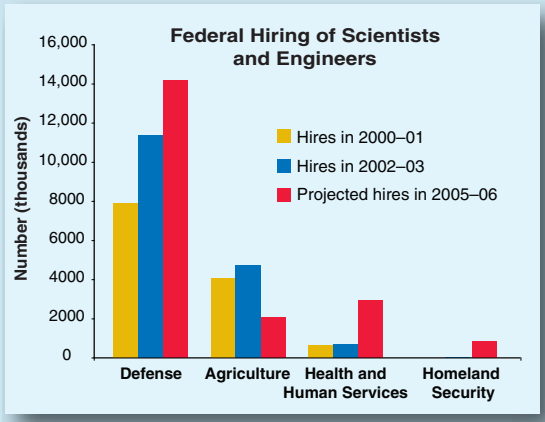
Washington, D.C., represents the first major hire for NSF Director Arden Bement.

Lightfoot says he's keen to expand NSF's collaborations abroad, in line with graduate training alliances he created at Georgetown involving Mexico and China. He also hopes to strengthen interagency cooperation, citing opportunities to support areas of basic research in neuroscience that could be jettisoned by the National Institute of Mental Health (*Science*, 22 October 2004, p. 602). Lightfoot joins NSF on 1 June, some 15 months after the departure of Norman Bradburn.

Got any tips for this page? E-mail people@aaas.org

DATA POINT

Job openings. The U.S. government expects to hire nearly 24,000 scientists and engineers over the next 2 years, according to a report released last week by the nonprofit Partnership for Public Service and the National Academy of Public Administration. The Department of Defense will continue to be the biggest government employer of scientific talent, increasing the number of new hires from 11,368 in 2002 and 2003 to a projected 14,156. The report cites national security concerns as one driving force. The Department of Agriculture, for example, plans to hire 400 new specialists to work on food inspections.



PIONEERS

Fighting inequity. Jean Fort kept working on one issue even after she retired last year as assistant vice chancellor for research at the University of California (UC), San Diego. Last month that work paid off, in the form of uniform health insurance for the 6000 postdoctoral scholars across UC's 10 campuses.

The new health plan is the first of its kind in the nation, and it puts an end to disparities in coverage based on the postdoc's funding source. It's the product of a 6-year effort that replaced a tangle of ad hoc procedures on the different campuses with a single, consistent set of rules (see nextwave.sciencemag.org/cgi/content/full/2005/01/06/9).

Fort gives credit to scores of fellow administrators who worked to change decades-old practices. But Fort was the "locomotive," says Sam Castaneda, director of Berkeley's Visiting Scholar and Postdoctoral Affairs Program: "If it weren't for her, it never would have rolled out so quickly or smoothly."

DEATHS

Ageless. Biologist Ernst Mayr died in Bedford, Massachusetts, on 3 February, leaving behind a legacy of 700 research papers and more than two dozen books that have shaped evolutionary thinking. He celebrated his 100th birthday last year.

An ornithologist turned evolutionary biologist who began his academic career at the American Museum of Natural History and later taught at Harvard University, Mayr's signature contribution to the field was to show that new species do arise from isolated populations as Charles Darwin had thought. In 1942 he published a groundbreaking book, *Systematics and the Origin of Species*, that gave a 20th-century perspective on Darwin's ideas. He continued to publish well after his retirement from Harvard in 1975: Last year, Cambridge University came out with his 25th book titled *What Makes Biology Unique*. In recent months, he was in the midst of a comprehensive reanalysis of Darwin's ideas.

Mayr not only shaped scientific history but also served as an outstanding historian of science, says Harvard's James Hanken. In addition, "he was a tremendous champion of natural history museums."

THEY SAID IT

"One of the things we tried to do with this film was to show what scientists are really like. ... They're not driven by a materialistic value system. They're seeking something else, something more important."

—James Cameron on his latest offering, titled *Aliens of the Deep*, in a recent interview to *The New York Times*. The film is a 48-minute IMAX feature in which Cameron explores the deep sea with a crew of young marine

Published by AAAS

Letters to the Editor

Letters (~300 words) discuss material published in *Science* in the previous 6 months or issues of general interest. They can be submitted through the Web (www.submit2science.org) or by regular mail (1200 New York Ave., NW, Washington, DC 20005, USA). Letters are not acknowledged upon receipt, nor are authors generally consulted before publication. Whether published in full or in part, letters are subject to editing for clarity and space.

Questions About *Orrorin* Femur

THE RECENT PUBLICATION OF CT SCAN DATA from the Lukeino Upper Miocene proximal femur assigned to "*Orrorin tugenensis*" ("External and internal morphology of the BAR 1002′00 *Orrorin tugenensis* femur," K. Galik *et al.*, Reports, 3 Sept. 2004, p. 1450) was surprising. The paper shows ingenuity and skill in correcting for improper specimen orientation in the original scans published in 2002 (*1*). However, the adjusted results are still inadequate to determine whether the femoral neck's cortical thickness conforms to a human, intermediate, or chimpanzee pattern.

The 6-million-year-old femur in question.

CREDIT: MARC DEVILLE/GAMMA

When found, the fossil's femoral neck was naturally fractured (and then glued) at essentially the exact location most needed for an accurate analysis. The explanation by one of the authors that "the bone was broken in a zigzag pattern that made it difficult to photograph" ("Oldest human femur wades into controversy," A. Gibbons, News of the Week, 24 Sept. 2004, p. 1885) is unsatisfactory. Measurements of actual cortical thickness should have been made before the bone was glued together. The bone should be unglued to extract essential photographic and metric data.

Mineralization of matrix within the neck may have affected the CT results. Conventional x-rays of the specimen should be published, particularly as such images are often more illuminating than CT scans (*2*).

We agree that the Lukeino femur's external morphology suggests some form of bipedality. Yet the more detailed original scans (*1*) appear to show a distinct superior cortex different from *Australopithecus* and humans, with the cortex distribution being more primitive than that seen in any other hominid, including *Australopithecus* (a point that Galik *et al*. concede in their Report). However, authors B. Senut and M. Pickford have claimed (*1*, *3*), and apparently continue to assert (24 Sept. 2004, p. 1885), that this femur is so derived as to exclude *Australopithecus* from direct human ancestry.

Given the importance of the Lukeino femur, we urge its discoverers to make available evidence to support their assertions. The required evidence is (i) photographs, measurements, and drawings of its broken neck; (ii) conventional anteroposterior x-rays; and (iii) higher-resolution CT scans obtained with proper femoral orientation. Exceptional claims demand exceptional evidence; the adjustment of previously published data does not suffice.

JAMES C. OHMAN,[1] C. OWEN LOVEJOY,[2] TIM D. WHITE[3]

[1]Research Centre in Evolutionary Anthropology and Palaeoecology, School of Biological and Earth Sciences, Liverpool John Moores University, Liverpool L3 3AF, UK. E-mail: J.C.Ohman@ livjm.ac.uk. [2]Department of Anthropology and Division of Biomedical Sciences, Kent State University, Kent, OH 44242, USA. E-mail: olovejoy@aol.com. [3]Department of Integrative Biology and Human Evolution Research Center, University of California at Berkeley, Berkeley, CA 94720, USA. E-mail: timwhite@socrates.berkeley.edu

References

1. M. Pickford, B. Senut, D. Gommery, J. Treil, *C. R. Palevol.* **1**, 191 (2002).
2. C. O. Lovejoy, R. S. Meindl, J. C. Ohman, K. G. Heiple, T. D. White, *Am. J. Phys. Anthropol.* **119**, 97 (2002).
3. B. Senut *et al.*, *C. R. Earth Planet. Sci.* **332**, 137 (2001).

Response

WE APPRECIATE *SCIENCE*'S CONTINUING coverage of the Tugen Hills fossils, including our Report, the subsequent news story (A. Gibbons, "Oldest human femur wades into controversy," 24 Sept. 2004, p. 1885), and the Letter by Ohman *et al.* This discussion documents the broadening consensus that *Orrorin* was bipedal, a recent shift; previously, some palaeoanthropologists expressed marked skepticism on this point (*1*), which no longer seems so exceptional given its acceptance by Ohman *et al*. Now criticism has shifted to the quality of the CT scans on which our analyses were based. In assessing these, we observed scrupulously the Feynman principle: "If you are doing an experiment, you should report everything that you think might make it invalid—not only what you think is right about it." We published Hounsfield plots that showed imperfect resolution of the scans, also noted in the legends to figs. 2 and 3.

The specific points raised by Ohman *et al.* fall into two categories: technical questions and phylogenetic speculations. Regarding technical matters, it is our understanding that the initial studies were carried out under serious constraints of time and other resources. In any case, our Report included 10 times as many CT scans as had been available previously, and we already have made it clear that we plan to rescan and study the existing fossils if funds are made available.

As far as phylogenetic speculations, a fuller understanding of the first several million years of human ancestry awaits the outcome of studies (already under way by other members of our research group) of the equivocal hominoid remains from Chad (*2*), as well as some much more comprehensive results from the by now decade-long analysis of the *Ardipithecus* (née *Australopithecus*) *ramidus* fossils (*3*), the reported fragility of which nonetheless should not preclude the making of CT scans and publication of what they show.

ROBERT B. ECKHARDT,[1] KAROL GALIK,[2] ADAM J. KUPERAVAGE[1]

[1]Laboratory for the Comparative Study of Morphology, Mechanics and Molecules, Department of Kinesiology, The Pennsylvania State University, University Park, PA 16802, USA. [2]Orthopedic Biomechanics Laboratory, Allegheny General Hospital, Pittsburgh, PA 15212, USA.

References

1. A. Gibbons, *Science* **295**, 1214 (2002).
2. M. Brunet *et al.*, *Nature* **418**, 145 (2002).
3. T. D. White, G. Suwa, B. Asfaw, *Nature* **371**, 306 (1994).

FBI mtDNA Database: A Cogent Perspective

IN THEIR LETTER "PROBLEMS IN FBI MTDNA database" (3 Sept. 2004, p. 1402), H.-J. Bandelt *et al.* state that "the U.S. National Institute of Justice has regularly resisted comprehensive evaluations of the science underlying forensic techniques" and that a possible answer for this "can be found in the poor quality of the forensic human mitochondrial DNA (mtDNA) database used by the Federal Bureau of Investigation (FBI)." Such hyperbole is unsubstantiated and misdirects the reader from the important issues regarding the forensic human mitochondrial DNA (mtDNA) database used by the FBI: What quality practices are in place, have the detected errors been corrected, and most importantly, if a few errors still reside in the database, what is the practical impact on estimating the rarity of a mtDNA profile?

We are perplexed by the inference of resistance. The National Institute of Justice has convened a multiyear Commission on DNA Typing (*1*). Also, the National Academy of Sciences has reviewed forensic DNA typing (*2*), and there are established quality assurance standards, audit documents and practices, proficiency testing requirements, and mtDNA interpretation guidelines (*3*).

It is well known that most, if not all, mtDNA databases contain some sequences that are in error (*4*), mostly due to human error. The errors within the Scientific Working Group on DNA Analysis Methods (SWGDAM) database have been and continue to be addressed (*5*, *6*). Our forensic mtDNA database is one of the first to be placed on the Web (*7*), so the data could be reviewed, potential errors identified, and those verified errors corrected.

Bandelt *et al.* identified eight samples that belong to Haplogroup A that could be in error in our database (*5*). Three were in error at one position out of 600 bases. The other five were not in error, but are true reversals. Most, if not all, of the samples described by Bandelt *et al.* and us (*5*, *8*) had been corrected before publication of their Letter. Our database is not static. The implications by Bandelt *et al.* of "poor quality" are exaggerated.

Bandelt *et al.* do not describe the impact of the errors on estimating the rarity of an evidentiary mtDNA sequence (*9*). They do not address the intra-individual variation in the population(s), nor the estimate of the rarity of the profile when there are errors in some sequences. Budowle *et al.* (*5*) have reported such an analysis. Although we advocate correcting errors, the impact on forensic estimates is nominal or is not different before and after correcting the samples in error.

Bandelt *et al.* suggest that "biochemical problems are manifest" because a few samples have undetermined nucleotides (labeled as N's). If an unresolved ambiguity is observed at any site, the base state is listed as an "N" (*10*). Again, Bandelt *et al.* do not address the impact of an N designation at a site on mtDNA profile frequency estimates. N's are treated as "wild cards." Therefore, the number of profiles (including those with N's) in the SWGDAM database that match the evidence profile will either be equal to the number of profiles that should match or be greater.

Reviews continue to increase the quality of the SWGDAM mtDNA population database. The forensic community can have confidence in its reliability.

BRUCE BUDOWLE AND DEBORAH POLANSKEY
FBI Laboratory, 2501 Investigation Parkway, Quantico, VA 22135, USA.

References
1. See www.ojp.usdoj.gov/nij/dna/.
2. National Research Council, *The Evaluation of Forensic DNA Evidence* (National Academies Press, Washington, DC, 1996).
3. "Quality assurance standards for forensic DNA testing laboratories," *Forensic Sci. Commun.* **2** (no. 3) (July 2000) (available at www.fbi.gov/hq/lab/fsc/backissu/july2000/codispre.htm).
4. Y.-G. Yao *et al.*, *Forensic Sci. Int.* **141**, 1 (2004).
5. B. Budowle *et al.*, *J. Forensic Sci.* **49**, 1256 (2004).
6. D. Polanskey, CODISmt v1.2 mtDNA Database update, presented at American Academy of Forensic Sciences Meeting, Dallas, TX, 2004.
7. K. L. Monson *et al.*, *Forensic Sci. Commun.* **4** (no. 2) (April 2002) (available at www.fbi.gov/hq/lab/fsc/backissu/april2002/miller1.htm).
8. D. Polanskey *et al.*, *Forensic Sci. Commun.* **7** (no. 1) (Jan. 2005) (available at www.fbi.gov/hq/lab/fsc/current/research/2005research.htm).
9. B. Budowle *et al.*, *Forensic Sci. Int.* **103**, 23 (1999).
10. J. O'Callaghan *et al.*, presentation at the 15th International Symposium on Human Identification, Phoenix, AZ, Oct. 2004.

Brazil's Nuclear Activities

THE POLICY FORUM "BRAZIL'S NUCLEAR PUZZLE" by L. Palmer and G. Milhollin (22 Oct. 2004, p. 617) contains preposterous allegations about my country's nuclear activities.

Palmer and Milhollin state that "Resende [a new Brazilian nuclear facility] will have the

potential to produce enough ^{235}U to make five to six implosion-type warheads per year." This assertion attempts to sell purely speculative data as hard fact and fails to point out that high levels of enrichment demand specifically designed installations.

The authors write that "Brazil took the extraordinary step of barring the plant's doors to the IAEA's [International Atomic Energy Agency's] inspectors." Every nuclear facility in Brazil is under IAEA safeguards, and IAEA inspectors have already had access to Resende's cascade hall on seven occasions.

Palmer and Milhollin state that the physical screen Brazil has built around its centrifuges at Resende "will make it harder—if not impossible—for the IAEA to do its job." Brazil and the IAEA are working for the placement of effective and utterly credible safeguards. The IAEA will have access to all tubes, valves, and connections in the cascade to ensure that no uranium is being enriched beyond the 5% level.

Contrary to Palmer and Milhollin's assertions, Brazil is not a "serious challenge to the IAEA's authority." Brazil has an impeccable relationship with the IAEA and is simply discussing with them how to reconcile Brazil's commitments to the IAEA with the country's legitimate right to protect proprietary technology.

Palmer and Milhollin claim that "[d]uring the 1980s, Brazil ran a secret effort to build an atomic bomb…" In the 1980s, Brazil implemented efforts to develop an autonomous capacity of enriching uranium for the production of electricity.

Palmer and Milhollin suggest that "the United States [should] convince Brazil to… be a good nuclear citizen." For decades, Brazil has been a committed champion of the twin causes of disarmament and nonproliferation and is party to all major treaties and instruments (NPT, Tlatelolco, CTBT, NSG, and MTCR, among others). With the creation of ABACC, Brazil and Argentina pioneered a scheme for bilateral nuclear inspections that sets an example for other regions. Brazil is therefore a model nuclear citizen.

ROBERTO ABDENUR
Brazilian Ambassador to the United States, Brazilian Embassy, 3006 Massachusetts Avenue, NW, Washington, DC 20008, USA.

Response

THE BRAZILIAN AMBASSADOR'S RESPONSE TO our Policy Forum is gravely misinformed.

He claims that during the 1980s, Brazil sought to enrich uranium for the production of electricity, implying that Brazil did not seek to build a nuclear weapon. In doing so, the Ambassador is in denial about something his own government and the rest of the world acknowledged in the early 1990s (*1*).

In noting that the IAEA has had access to the Resende facility on several occasions, the Ambassador is glossing over the fact that, for several months earlier this year, IAEA inspectors and Brazilian officials were at an impasse regarding inspections. In violation of its obligations under the Nuclear Nonproliferation Treaty, Brazil had refused to allow IAEA inspectors to view its centrifuges. Instead, Brazil erected a screen around the centrifuges, citing concerns that the IAEA would not protect its technology. Denying access to the facility in violation of treaty obligations and under the unlikely suggestion that the IAEA cannot protect trade secrets can only be viewed as a challenge to the IAEA's authority.

The Ambassador further misses the point when he states that specifically designed installations are required to produce high-enriched uranium. Our point is that, if the centrifuges are hidden from the inspectors' view, Brazil could be siphoning off low-enriched uranium that would allow it to stockpile this material. If Brazil decided to become a nuclear power, such a

stockpile would make it easier for Brazil to build a bomb before the world could react. Although we do not assert that Brazil aims to produce nuclear weapons, the capability would be there.

Since our Policy Forum was published in October, the IAEA and Brazil have reportedly come to an agreement whereby IAEA inspectors will be able to view the valves and tubes leading to and from the centrifuges.

LIZ PALMER AND GARY MILHOLLIN
Wisconsin Project on Nuclear Arms Control, Washington, DC 20006, USA.

Reference

1. J. Brooke, *N.Y. Times*, 9 Oct. 1990, p. A1.

Optimism About String Theory

THE NEWS FOCUS ARTICLE "STRING THEORY gets real—sort of" (A. Cho, 26 Nov. 2004, p. 1460) nicely describes the historical issues and separation of string theory and phenomenology, but it does not convey the recent excitement and increased activity of the emerging area of "string phenomenology." The best documentation for that is people voting with their "feet"—last summer witnessed the third international string phenomenology meeting (in Ann Arbor), with the fourth already scheduled for Munich next summer, plus the first ever string phenomenology workshop, on which the article reported. String theory is exciting not only as a quantum theory of gravity, but because it allows us to address as research topics basic questions previously out of reach. Increasingly, many people are working in these areas and unifying different areas (such as collider physics, inflation, dark matter, quark masses, neutrino masses, and more) that could only be connected by a deep underlying theory. Whether such efforts are successful remains to be seen, but a number of people are more optimistic than they have been for a long time.

GORDON KANE
Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA.

CORRECTIONS AND CLARIFICATIONS

Reports: "Integration of visual and linguistic information in spoken language comprehension" by M. K. Tanenhaus *et al.* (16 June 1995, p. 1632). The paper should have included the following reference to an earlier description of eye movement recording used to track rapid mental processes: R. M. Cooper, The control of eye fixation by the meaning of spoken language, *Cognit. Psychol.* **6**, 84 (1974). The authors regret the omission.

PHYSICS

Treks of Imagination

Frank Wilczek

Roger Penrose is a highly original and creative mathematical physicist. He pioneered the development of global methods in general relativity, which transformed and greatly deepened our understanding of black holes, revealing them to be dynamic entities rather than dead, frozen ashes. Overturning long-accepted dogmas, he also discovered the possibility of quasiperiodic tilings, which have turned out to describe actual materials (quasicrystals).

The Road to Reality
A Complete Guide to the Laws of the Universe
by Roger Penrose

Jonathan Cape, London, 2004. 1122 pp. £30, C$85. ISBN 0-224-04447-8. Knopf, New York, 2005. $40. ISBN 0-679-45443-8.

In recent years, Penrose has become a successful author of popular books on science. In this too, his approach has been highly original. *The Emperor's New Mind* (*1*) and *Shadows of the Mind* (*2*) are long, demanding books that include brilliant and uncompromising exposition of difficult material. For example, they provide honest, self-contained accounts of universal Turing machines, Gödel's theorem, and the foundations of quantum mechanics that are models of lucid exposition. But these books also make connections and propound scientific theses that are, to say the least, controversial. They culminate in claims that quantum coherence and quantum gravity are implicated in normal brain function and consciousness, claims that have not won wide acceptance.

The Road to Reality resembles those earlier books in its eclectic style, and it includes consideration of some of the same themes. But it comes at them from a different perspective and develops them in quite different directions. Here the emphasis is on physics, not biology.

The first thing to say about *The Road to Reality* is that it is a big, sprawling book. As the text progresses, the level of sophistication expected of the reader ascends from modest beginnings to truly dizzying heights. The exposition might be self-contained at some formal level, but in any realistic sense it is not. I don't imagine that there are many readers who will first learn from the book both what a complex number is and what a holomorphic line bundle on twistor space is.

Nor need they try. For despite the title, the second thing to say about what's inside the book is that Penrose does not present anything like a well-ordered, sequential path leading to the Holy Grail "reality." Rather, one finds a series of intellectual treks, toward the end quite strenuous, with a brilliant and engaging, if not entirely reliable, guide into some wild frontiers of fundamental physics and cosmology. The Grail is never espied; too bad. But because readers are liberated from the obligation to follow a single road, they can feel free to pick and choose from among the treks on offer—and to turn back, without a deep sense of loss, when the going gets too rough.

Because *The Road to Reality* is so diverse and multi-tiered, I find it useful to discuss the book from three different perspectives: as a survey for novices, as stimulation for sophisticates, and as a scientific treatise.

Dante astray in the Dusky Woods. The first of Gustave Doré's illustrations (1861) of Dante's *Inferno*.

I imagine that a bright and ambitious teenager contemplating a career in mathematics or physics would find much to enjoy and to savor in the book. Its first half (running for over 500 pages) presents a smorgasbord of essentially mathematical ideas, including unusual perspectives on complex variables, Fourier analysis, and connections between these that motivate the concept of hyperfunctions. The discussions of the conformal geometry of special relativity and of spinors are real gems. This part would suit beginners well, as a stimulus to their imaginations. The second half of the book enters into essentially physical material. Here, I think failure to ground the discussion adequately in empirical facts renders it quite unsuitable for those new to the field. They will not obtain from the book a sound knowledge of the basics of the subjects under discussion, nor will they be in a position to judge the relative credibility of conventional ideas and Penrose's alternatives. It is as if in a trial one were exposed only to the prosecution's summation, without the presentation of physical evidence or the testimony of witnesses (much less the defense's arguments).

Sophisticated physicists will pass over the book's first half rapidly, enjoying the gems. In the second half, they will find several provocative ideas. Specifically, Penrose argues that there must be a genuine physical process, involving non-unitary evolution, that implements the "collapse of the wave function" in quantum theory. He speculates that this "R-Process" will emerge from a quantum theory of gravity. He also argues forcefully that the standard initial state assumed in big bang cosmology, which posits thermal equilibrium for matter but near-perfect order in the gravity field (i.e., uniformity), is highly unlikely—in the precise sense of having extremely low entropy relative to the maximum available. He speculates that an intrinsic time asymmetry of physical law, perhaps related to the R-Process, will explain this anomaly. He expresses at length his frustration with string theory and sketches loop variable, spin network, and twistor alternatives. The last of these I found particularly interesting, in its introduction of beautiful mathematics and a controlled form

The reviewer is in the Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139–4307, USA. E-mail: wilczek@mit.edu

Published by AAAS

of non-locality, but at present twistor ideas appear more as the desire for a physical theory than the embodiment of one.

Regarded as a scientific treatise, *The Road to Reality* is in many ways problematic. By nominally addressing a substantive discussion of frontier issues in theoretical physics and cosmology to a popular audience, an author deprives himself of the discipline of having to provide details, to address concrete experimental issues, or to pitch the level of his argumentation to peers capable of judging them critically. Galileo pulled this off brilliantly, but times were much simpler then! The worst parts of the book are the chapters on high-energy physics and quantum field theory, which in spite of their brevity contain several serious blunders: The Cabibbo angle does not govern the mixing of K^0 and $\bar{K}^0$ mesons to make the long- and short-lived Ks. There are not alternative directions of electroweak symmetry breaking. And no associated disorder arises at that symmetry-breaking transition, any more than at the analogous transition in ordinary superconductors.

To summarize, there's much to admire and profit from in this remarkable book, but judged by the highest standards *The Road to Reality* is deeply flawed.

References

1. R. Penrose, *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics* (Oxford Univ. Press, Oxford, 1989).
2. R. Penrose, *Shadows of the Mind: An Approach to the Missing Science of Consciousness* (Oxford Univ. Press, Oxford, 1994).

10.1126/science.1106081

A DAY OUT: BIOGRAPHY

Refuge from Berlin's Bustle

John Bohannon

"The sailing ship, the distant view, the lonely walks in autumn, the relative silence, it is paradise." This is how Albert Einstein described his summer home just outside Berlin in the village of Caputh, where he lived from 1929 to 1932. He could not have known these would be the last carefree years of his life. When the Nazis seized power in 1933, he was visiting the United States and he settled there. Soon, helpless, he would witness his wife Elsa's death from a painful illness and later the application of his revolutionary theories in the creation of the most destructive weapons ever known.

The reviewer is at Choriner Strasse 74, 10119 Berlin, Germany. Web site: www.johnbohannon.org

CREDIT: STIFTUNG ARCHIV AKADEMIE DER KÜNSTE, SAMMLUNG BAUKUNST, BERLIN

Einstein's Summer House in Caputh
Einstein Forum, Am Neuen Markt 7, 14467 Potsdam, Germany. www.einsteinforum.de

Einstein is receiving even greater attention than usual in this pleasingly symmetric year, 50 years after his death and 100 years after the publication of his world-changing trio of publications on the quantum theory of light, Brownian motion, and special relativity. For those seeking a more private glimpse into his life, a trip to Berlin would be timely. In preparation for visitors, the Einstein Forum (an interdisciplinary institution formed in 1993 to promote innovative thinking and engage the public) is renovating the Caputh summer house and plans to offer tours starting in May.

Walking along the damp path through the birch trees, slipping through the back door of the house, and opening the wide windows to gaze out over the gracious curve of Templiner Lake, one immediately sees what Einstein meant by paradise. The house was an escape from the intruding outer world.

By 1929, Einstein was already a household name. Ten years earlier, his general theory of relativity had been triumphantly confirmed by the observation of starlight bending around the eclipsed sun, and he had received the 1921 Nobel Prize in physics (although the citation mentioned the photoelectric effect, not relativity). As word spread of his 50th birthday, Einstein was inundated with letters and presents from around the world. But what he most wanted was a quiet refuge where he could entertain friends, spend time with his wife and two stepdaughters, and think about the unified field theory, which he would pursue for the rest of his life and which still eludes physicists today. And so, using most of his savings to buy the land and build a house, he got his retreat.

Einstein's desire for a wooden house attracted the architect Konrad Wachsmann, who designed the block house at Caputh and became a dear friend. Acting from Princeton, Einstein later helped him flee Nazi Germany for the United States. There, in the 1940s, Wachsmann worked with Walter Gropius. Together they developed a system for producing prefabricated wood houses that would gain him an international reputation and help radically alter the suburban landscape. As one of his earliest designs using wood, the Caputh house has an added historical significance.

As Einstein no doubt would have wanted, the house has not been turned into a shrine. Instead, it continues to be used for the annual Nobel lectures held by the Forum and for academic retreats. Rather than rummaging for furniture in antique shops to match the original contents (which were lost in the years of first Nazi and then communist East German control), the Forum is fitting the house with functional, tasteful equivalents. As the Forum's Rüdiger Zill puts it, this is the "honest" approach.

A 15-minute drive away in Potsdam, black-and-white photographs from Einstein's Caputh years are on display at the Einstein Forum. One can't help but smile seeing these images; some are iconic, such as Einstein setting off in Tümmler, the beloved sailboat he kept moored on the Templiner. But many others are intimate and spontaneous: His stepdaughters recline in sunchairs with obvious pleasure. Einstein emerges serenely from the door in rumply clothes or gazes out the window with a look of utter peace. These reveal Einstein at his most unguarded and, perhaps, optimistic.

10.1126/science.1110157

BROWSINGS

Einstein 1905. The Standard of Greatness. *John S. Rigden.* Harvard University Press, Cambridge, MA, 2005. 185 pp. $21.95, £14.95. ISBN 0-674-01544-4.

Between March and September 1905, Einstein wrote five *Annalen der Physik* papers that would greatly influence 20th-century physics. These present the argument, from considerations of entropy, that light consists of quanta; Einstein's dissertation on the determination of molecular dimensions; his theory of Brownian motion; the theory of special relativity; and the derivation of $m = E/c^2$. For each paper, Rigden discusses the background, underlying ideas, content, and organization before surveying its reception and impact. General readers who wish to understand the magnitude of what Einstein accomplished during his *annus mirabilis* will find this lucid, nonmathematical account ideal.

of non-locality, but at present twistor ideas appear more as the desire for a physical theory than the embodiment of one.

Regarded as a scientific treatise, *The Road to Reality* is in many ways problematic. By nominally addressing a substantive discussion of frontier issues in theoretical physics and cosmology to a popular audience, an author deprives himself of the discipline of having to provide details, to address concrete experimental issues, or to pitch the level of his argumentation to peers capable of judging them critically. Galileo pulled this off brilliantly, but times were much simpler then! The worst parts of the book are the chapters on high-energy physics and quantum field theory, which in spite of their brevity contain several serious blunders: The Cabibbo angle does not govern the mixing of K^0 and $\bar{K}^0$ mesons to make the long- and short-lived Ks. There are not alternative directions of electroweak symmetry breaking. And no associated disorder arises at that symmetry-breaking transition, any more than at the analogous transition in ordinary superconductors.

To summarize, there's much to admire and profit from in this remarkable book, but judged by the highest standards *The Road to Reality* is deeply flawed.

References

1. R. Penrose, *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics* (Oxford Univ. Press, Oxford, 1989).
2. R. Penrose, *Shadows of the Mind: An Approach to the Missing Science of Consciousness* (Oxford Univ. Press, Oxford, 1994).

10.1126/science.1106081

A DAY OUT: BIOGRAPHY

Refuge from Berlin's Bustle

John Bohannon

"The sailing ship, the distant view, the lonely walks in autumn, the relative silence, it is paradise." This is how Albert Einstein described his summer home just outside Berlin in the village of Caputh, where he lived from 1929 to 1932. He could not have known these would be the last carefree years of his life. When the Nazis seized power in 1933, he was visiting the United States and he settled there. Soon, helpless, he would witness his wife Elsa's death from a painful illness and later the application of his revolutionary theories in the creation of the most destructive weapons ever known.

The reviewer is at Choriner Strasse 74, 10119 Berlin, Germany. Web site: www.johnbohannon.org

CREDIT: STIFTUNG ARCHIV AKADEMIE DER KÜNSTE, SAMMLUNG BAUKUNST, BERLIN

Einstein's Summer House in Caputh
Einstein Forum, Am Neuen Markt 7, 14467 Potsdam, Germany. www.einsteinforum.de

Einstein is receiving even greater attention than usual in this pleasingly symmetric year, 50 years after his death and 100 years after the publication of his world-changing trio of publications on the quantum theory of light, Brownian motion, and special relativity. For those seeking a more private glimpse into his life, a trip to Berlin would be timely. In preparation for visitors, the Einstein Forum (an interdisciplinary institution formed in 1993 to promote innovative thinking and engage the public) is renovating the Caputh summer house and plans to offer tours starting in May.

Walking along the damp path through the birch trees, slipping through the back door of the house, and opening the wide windows to gaze out over the gracious curve of Templiner Lake, one immediately sees what Einstein meant by paradise. The house was an escape from the intruding outer world.

By 1929, Einstein was already a household name. Ten years earlier, his general theory of relativity had been triumphantly confirmed by the observation of starlight bending around the eclipsed sun, and he had received the 1921 Nobel Prize in physics (although the citation mentioned the photoelectric effect, not relativity). As word spread of his 50th birthday, Einstein was inundated with letters and presents from around the world. But what he most wanted was a quiet refuge where he could entertain friends, spend time with his wife and two stepdaughters, and think about the unified field theory, which he would pursue for the rest of his life and which still eludes physicists today. And so, using most of his savings to buy the land and build a house, he got his retreat.

Einstein's desire for a wooden house attracted the architect Konrad Wachsmann, who designed the block house at Caputh and became a dear friend. Acting from Princeton, Einstein later helped him flee Nazi Germany for the United States. There, in the 1940s, Wachsmann worked with Walter Gropius. Together they developed a system for producing prefabricated wood houses that would gain him an international reputation and help radically alter the suburban landscape. As one of his earliest designs using wood, the Caputh house has an added historical significance.

As Einstein no doubt would have wanted, the house has not been turned into a shrine. Instead, it continues to be used for the annual Nobel lectures held by the Forum and for academic retreats. Rather than rummaging for furniture in antique shops to match the original contents (which were lost in the years of first Nazi and then communist East German control), the Forum is fitting the house with functional, tasteful equivalents. As the Forum's Rüdiger Zill puts it, this is the "honest" approach.

A 15-minute drive away in Potsdam, black-and-white photographs from Einstein's Caputh years are on display at the Einstein Forum. One can't help but smile seeing these images; some are iconic, such as Einstein setting off in Tümmler, the beloved sailboat he kept moored on the Templiner. But many others are intimate and spontaneous: His stepdaughters recline in sunchairs with obvious pleasure. Einstein emerges serenely from the door in rumply clothes or gazes out the window with a look of utter peace. These reveal Einstein at his most unguarded and, perhaps, optimistic.

10.1126/science.1110157

BROWSINGS

Einstein 1905. The Standard of Greatness. *John S. Rigden.* Harvard University Press, Cambridge, MA, 2005. 185 pp. $21.95, £14.95. ISBN 0-674-01544-4.

Between March and September 1905, Einstein wrote five *Annalen der Physik* papers that would greatly influence 20th-century physics. These present the argument, from considerations of entropy, that light consists of quanta; Einstein's dissertation on the determination of molecular dimensions; his theory of Brownian motion; the theory of special relativity; and the derivation of $m = E/c^2$. For each paper, Rigden discusses the background, underlying ideas, content, and organization before surveying its reception and impact. General readers who wish to understand the magnitude of what Einstein accomplished during his *annus mirabilis* will find this lucid, nonmathematical account ideal.

of non-locality, but at present twistor ideas appear more as the desire for a physical theory than the embodiment of one.

Regarded as a scientific treatise, *The Road to Reality* is in many ways problematic. By nominally addressing a substantive discussion of frontier issues in theoretical physics and cosmology to a popular audience, an author deprives himself of the discipline of having to provide details, to address concrete experimental issues, or to pitch the level of his argumentation to peers capable of judging them critically. Galileo pulled this off brilliantly, but times were much simpler then! The worst parts of the book are the chapters on high-energy physics and quantum field theory, which in spite of their brevity contain several serious blunders: The Cabibbo angle does not govern the mixing of K^0 and $\bar{K}^0$ mesons to make the long- and short-lived Ks. There are not alternative directions of electroweak symmetry breaking. And no associated disorder arises at that symmetry-breaking transition, any more than at the analogous transition in ordinary superconductors.

To summarize, there's much to admire and profit from in this remarkable book, but judged by the highest standards *The Road to Reality* is deeply flawed.

References

1. R. Penrose, *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics* (Oxford Univ. Press, Oxford, 1989).
2. R. Penrose, *Shadows of the Mind: An Approach to the Missing Science of Consciousness* (Oxford Univ. Press, Oxford, 1994).

10.1126/science.1106081

A DAY OUT: BIOGRAPHY

Refuge from Berlin's Bustle

John Bohannon

"The sailing ship, the distant view, the lonely walks in autumn, the relative silence, it is paradise." This is how Albert Einstein described his summer home just outside Berlin in the village of Caputh, where he lived from 1929 to 1932. He could not have known these would be the last carefree years of his life. When the Nazis seized power in 1933, he was visiting the United States and he settled there. Soon, helpless, he would witness his wife Elsa's death from a painful illness and later the application of his revolutionary theories in the creation of the most destructive weapons ever known.

The reviewer is at Choriner Strasse 74, 10119 Berlin, Germany. Web site: www.johnbohannon.org

CREDIT: STIFTUNG ARCHIV AKADEMIE DER KÜNSTE, SAMMLUNG BAUKUNST, BERLIN

Einstein's Summer House in Caputh
Einstein Forum, Am Neuen Markt 7, 14467 Potsdam, Germany. www.einsteinforum.de

Einstein is receiving even greater attention than usual in this pleasingly symmetric year, 50 years after his death and 100 years after the publication of his world-changing trio of publications on the quantum theory of light, Brownian motion, and special relativity. For those seeking a more private glimpse into his life, a trip to Berlin would be timely. In preparation for visitors, the Einstein Forum (an interdisciplinary institution formed in 1993 to promote innovative thinking and engage the public) is renovating the Caputh summer house and plans to offer tours starting in May.

Walking along the damp path through the birch trees, slipping through the back door of the house, and opening the wide windows to gaze out over the gracious curve of Templiner Lake, one immediately sees what Einstein meant by paradise. The house was an escape from the intruding outer world.

By 1929, Einstein was already a household name. Ten years earlier, his general theory of relativity had been triumphantly confirmed by the observation of starlight bending around the eclipsed sun, and he had received the 1921 Nobel Prize in physics (although the citation mentioned the photoelectric effect, not relativity). As word spread of his 50th birthday, Einstein was inundated with letters and presents from around the world. But what he most wanted was a quiet refuge where he could entertain friends, spend time with his wife and two stepdaughters, and think about the unified field theory, which he would pursue for the rest of his life and which still eludes physicists today. And so, using most of his savings to buy the land and build a house, he got his retreat.

Einstein's desire for a wooden house attracted the architect Konrad Wachsmann, who designed the block house at Caputh and became a dear friend. Acting from Princeton, Einstein later helped him flee Nazi Germany for the United States. There, in the 1940s, Wachsmann worked with Walter Gropius. Together they developed a system for producing prefabricated wood houses that would gain him an international reputation and help radically alter the suburban landscape. As one of his earliest designs using wood, the Caputh house has an added historical significance.

As Einstein no doubt would have wanted, the house has not been turned into a shrine. Instead, it continues to be used for the annual Nobel lectures held by the Forum and for academic retreats. Rather than rummaging for furniture in antique shops to match the original contents (which were lost in the years of first Nazi and then communist East German control), the Forum is fitting the house with functional, tasteful equivalents. As the Forum's Rüdiger Zill puts it, this is the "honest" approach.

A 15-minute drive away in Potsdam, black-and-white photographs from Einstein's Caputh years are on display at the Einstein Forum. One can't help but smile seeing these images; some are iconic, such as Einstein setting off in Tümmler, the beloved sailboat he kept moored on the Templiner. But many others are intimate and spontaneous: His stepdaughters recline in sunchairs with obvious pleasure. Einstein emerges serenely from the door in rumply clothes or gazes out the window with a look of utter peace. These reveal Einstein at his most unguarded and, perhaps, optimistic.

10.1126/science.1110157

BROWSINGS

Einstein 1905. The Standard of Greatness. *John S. Rigden.* Harvard University Press, Cambridge, MA, 2005. 185 pp. $21.95, £14.95. ISBN 0-674-01544-4.

Between March and September 1905, Einstein wrote five *Annalen der Physik* papers that would greatly influence 20th-century physics. These present the argument, from considerations of entropy, that light consists of quanta; Einstein's dissertation on the determination of molecular dimensions; his theory of Brownian motion; the theory of special relativity; and the derivation of $m = E/c^2$. For each paper, Rigden discusses the background, underlying ideas, content, and organization before surveying its reception and impact. General readers who wish to understand the magnitude of what Einstein accomplished during his *annus mirabilis* will find this lucid, nonmathematical account ideal.

ETHICS

Forbidden Knowledge

Joanna Kempner,[1] Clifford S. Perlis,[2] Jon F. Merz[3]*

There is growing concern about the politicization and social control of science, constraining the conduct, funding, publication, and public use of scientific research (*1*). For example, human cloning and embryonic stem cell creation have been regulated or banned (*2*), activists have been lobbying Congress to remove funding from certain government-sponsored research (*3*–*5*), and science journal editors have been compelled to develop policies for publication of sensitive manuscripts (*6*, *7*).

Forbidden knowledge embodies the idea that there are things that we should not know (*8*–*15*). Knowledge may be forbidden because it can only be obtained through unacceptable means, such as human experiments conducted by the Nazis (*9*, *11*); knowledge may be considered too dangerous, as with weapons of mass destruction or research on sexual practices that undermine social norms (*8*, *9*, *12*); and knowledge may be prohibited by religious, moral, or secular authority, exemplified by human cloning (*10*, *12*).

Beyond anecdotal cases, little is known about what, and in what ways, science is constrained. To begin to fill this gap, we performed an interview study to examine how constraints affect what scientists do. In 2002–03, we conducted 10 pilot and 41 in-depth semistructured interviews with a sample of researchers drawn from prestigious U.S. academic departments of neuroscience, sociology, molecular and cellular biology, genetics, industrial psychology, drug and alcohol abuse, and computer science. We chose diverse disciplines to gauge the range, rather than prevalence, of experiences.

We asked subjects to consider their practices and rationales for limiting scientific inquiry or dissemination and to tell us about cases in which research in their own discipline had been constrained. Respondents reported a wide range of sensitive topics, including studies relating to human cloning, embryonic stem cells, weapons, race, intelligence, sexual behaviors, and addiction, as well as concerns about using humans and animals in research.

Nearly half the researchers felt constrained by explicit, formal controls, such as governmental regulations and guidelines codified by universities, professional societies, or journals. Respondents generally agreed that formal controls offered important protections. Less consensus surrounded the necessity, efficiency, or good sense of specific policies. Stem cell research was repeatedly identified as an example of an overly restricted area. Many respondents expressed a preference that scientists—not policy-makers—determine which research is too dangerous.

We were surprised, however, that respondents felt most affected by what we characterize as "informal constraints." Researchers sometimes only know that they have encountered forbidden knowledge when their research breaches an unspoken rule and is identified as problematic by legislators, news agencies, activists, editors, or peers. Studies by Kinsey *et al*. (*16*, *17*), Milgram (*18*), Humphreys (*19*), Herrnstein and Murray (*20*), and Rind *et al.* (*21*) were attacked only after publication. Many researchers (42%) described how their own work had been targeted for censure. One researcher was accused by activists of "murderous behavior" because he was incapable of reporting HIV+ subjects who admitted to unsafe sex practices in an anonymous survey. A sociologist published an article that undermined the central claim of a particular group, who allegedly then accused him of funding improprieties.

In other cases, the mere threat of social sanction deterred particular types of inquiry. Several researchers said that their choices to study yeast or mice instead of dogs were guided by fears of retribution from animal rights groups. As one respondent commented, "I would like to lunatic-proof my life as much as possible." Drug and alcohol researchers reported similar fears, stating that they had not pursued studies that might provoke moral outrage.

Finally, there may be unspoken rules shared by the community. As one respondent stated, "every microbiologist knows not to make a more virulent pathogen."

We failed to detect a coherent ethos regarding production of forbidden knowledge. Respondents at once decried external regulation and recognized the right of society to place limits on what and how science is done. They stated that scientists are "moral" and "responsible," but acknowledged cases in which scientists were sanctioned for acting outside the mainstream of their disciplines. They also said that, although information and "truth" had inherent utility, full and open publication was not always possible. Whereas most respondents worked hard to avoid controversy, others relished it.

In summary, formal and informal constraints have a palpable effect on what science is studied, how studies are performed, how data are interpreted, and how results are disseminated. Our results suggest that informal limitations are more prevalent and pervasive than formal constraints. Although formal constraints will bias science—by affecting what is studied and how it is studied—these biases are relatively transparent and amenable to political change. Informal constraints, in contrast, may be culturally ingrained and resistant to change, leaving few markers by which to assess their effects. We believe it is important to observe these constraints, assess their effects, and openly debate their desirability for science and society.

[1]School of Public Health, The University of Michigan, Ann Arbor, MI 48109–2029, USA. [2]Department of Dermatology, Brown University Medical School, Providence, RI 02903, USA. [3]Department of Medical Ethics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104–3308, USA.

*Author for correspondence. E-mail: merz@mail.med.upenn.edu

References and Notes

1. R. A. Charo, *J. Law Med. Ethics* **32**, 307 (2004).
2. G. Q. Daley, *New Engl. J. Med.* **349**, 211 (2003).
3. J. Kaiser, *Science* **300**, 403 (2003).
4. J. Kaiser, *Science* **302**, 758 (2003).
5. J. Kaiser, *Science* **302**, 966 (2003).
6. J. Couzin, *Science* **297**, 749 (2002).
7. Journal Editors and Authors Group, *Science* **299**, 1149 (2003).
8. C. Cohen, *New Engl. J. Med.* **296**, 1203 (1977).
9. D. Smith, *Hastings Center Rep.* **8** (6), 30 (1978).
10. G. Holton, R. S. Morison, Eds., *Limits of Scientific Inquiry* (Norton, New York, 1979).
11. D. Nelkin, in *Ethical Issues in Social Science Research*, T. L. Beauchamp, R. R. Faden, R. J. Wallace, L. Walters, Eds. (Johns Hopkins Univ. Press, Baltimore, MD, 1982), pp. 163–174.
12. R. Shattuck, *Forbidden Knowledge: From Prometheus to Pornography* (Harcourt Brace, New York, 1996).
13. D. B. Johnson, *Monist* **79**, 197 (1996).
14. B. Allen, *Monist* **79**, 294 (1996).
15. D. B. Johnson, *Sci. Eng. Ethics* **5**, 445 (1999).
16. A. C. Kinsey *et al.*, *Sexual Behavior in the Human Male* (Saunders, Philadelphia, 1948).
17. A. C. Kinsey *et al.*, *Sexual Behavior in the Human Female* (Saunders, Philadelphia, 1953).
18. S. Milgram, *Obedience to Authority: An Experimental View* (Harper Row, New York, 1974).
19. L. Humphreys, *Tearoom Trade: Impersonal Sex in Public Places* (Aldine, Chicago, 1970).
20. R. Herrnstein, C. Murray, *The Bell Curve: Intelligence and Class Structure in American Life* (Simon & Schuster, New York, 1996).
21. B. Rind *et al.*, *Psychol. Bull.* **124**, 22 (1998).
22. This study was approved by the University of Pennsylvania Institutional Review Board. We thank all respondents for their participation; B. Sitko for assistance; and C. Bosk, A. Caplan, J. Drury, C. Lee, and B. Sampat for comments. Supported by the Greenwall Foundation (J.K., C.S.P., J.F.M.) and the Robert Wood Johnson Foundation (J.K.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/307/5711/854/DC1

10.1126/science.1107576

ASTRONOMY

A Pulsar Bonanza

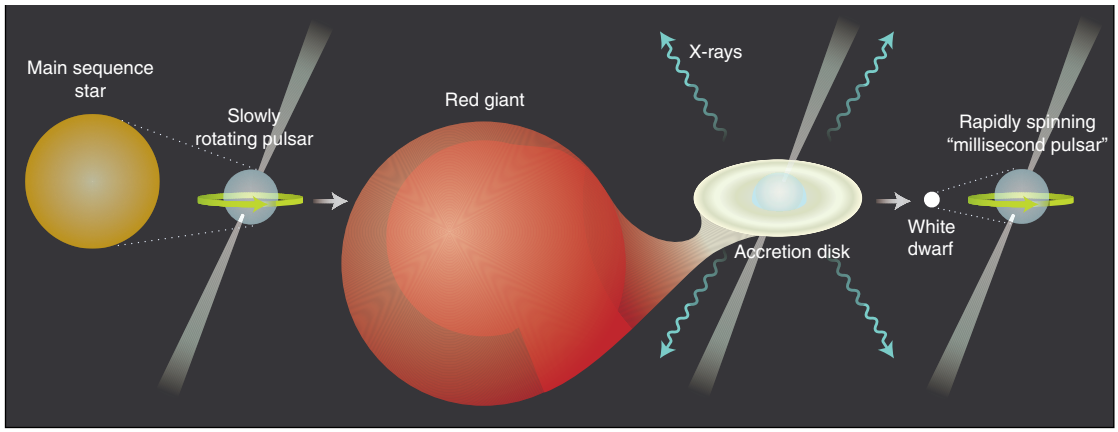
Duncan R. Lorimer

When a massive star explodes in a supernova, it leaves a compact object called a neutron star. The spin of the neutron star and its enormous magnetic field generate a rotating radiation beam. Rather like viewing a lighthouse from a distance, observers on Earth receive pulses of radiation each time the neutron star's beam crosses the line of sight of a radio telescope. The clocklike rotational stability of these objects—referred to as pulsars—has provided a wealth of insights into general relativity, galactic astronomy, planetary physics, and even cosmology.

Recent technological improvements and new instruments are enabling astronomers to find many new pulsars and to probe a rich variety of astrophysical settings. The latest breakthrough is the discovery of 21-millisecond pulsars in the globular cluster Terzan 5 by Ransom *et al.* [see page 892 of this issue (*1*)]. The study provides new opportunities for studying the extreme environments in globular clusters and the formation mechanisms for millisecond pulsars.

With rotation rates as fast as 642 Hz, millisecond pulsars act as cosmic flywheels that can sustain their rapid rotation on time scales of up to 10 billion years. In the most likely formation mechanism (see the first figure), a neutron star with a spin rate of a few hertz is spun up through the transfer of mass from a binary companion. The substantial heating of the material as it spirals into the strong gravitational field of the neutron star produces x-rays.

Such x-ray binaries exist in our own galaxy, but they are 10 times more common in globular clusters (dense conglomerations of 100,000 or more stars that are packed into a radius of about 10 light-years; they are about 200 times denser than the Milky Way). Globular clusters are excellent breeding grounds for millisecond pulsars, because the extremely high stellar density increases the probability of encounters between cluster members. For example, in a so-called exchange interaction (*2*), a neutron star or black hole can collide with a binary system and capture the more massive star to form a new binary.



Formation of a millisecond pulsar in a binary system. Initially, a slowly rotating pulsar orbits around a normal star (**left**). When the normal star reaches the end of its life and becomes a red giant, an x-ray binary is formed (**middle**). Material from the outer layers of the red giant is drawn toward the neutron star by its intense gravitational field, resulting in a disk of material that spirals into the neutron star. As a result of intense friction, the star is heated to more than 10^6 K and emits x-rays. The transfer of material onto the neutron star also transfers angular momentum from the orbit to the neutron star. This process can spin up the neutron star to rotation rates exceeding 500 Hz, consistent with the spin rates of the fastest rotating millisecond pulsars in Terzan 5. (**Right**) At the end of the red giant phase, the star sheds its outer layers and becomes a compact white dwarf that orbits around the rapidly rotating millisecond pulsar.

Before the new discoveries reported in (*1*), the record number of pulsars known in a single cluster was 22 in 47 Tucanae (*3*). Studies of the pulsars in 47 Tucanae, which have spin periods of 2 to 8 ms, have provided a wealth of information, including the first definite evidence for the existence of gas in a globular cluster (*4*).

As one of the densest and most massive clusters, Terzan 5 has long been thought to harbor many more pulsars than the three found in earlier searches (*5*–*7*). The main indication for this was the excess radio emission from the core of Terzan 5 (see the second figure). Assuming that the radio emission of the putative pulsar population is similar to that known from studies of other pulsars (*3*, *8*), this excess can be explained (*8*) by the combined emission of up to several hundred pulsars.

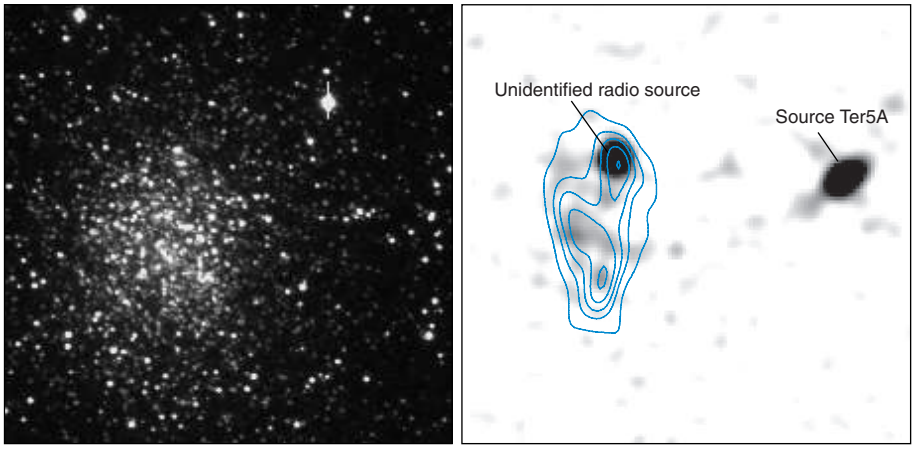
With the Green Bank Telescope, Ransom *et al.* (*1*) have carried out observations of Terzan 5 at 5 to 10 times higher sensitivity than previous searches (*5*–*7*) and have cracked the pulsar enigma in this cluster. With the 21 newly discovered pulsars, the total number of known pulsars in Terzan 5 is now 24. Individual pulsars are referred to by letters of the alphabet.

The spin periods and orbital parameters of the new pulsars are very different from those of the 47 Tucanae pulsars. The spin periods of the Terzan 5 pulsars span a much broader range (1.67 to 80 ms), including two rapidly spinning pulsars (O and P). Fourteen of the pulsars are binaries, including two systems with eccentric orbits (I and J). No such systems are known in 47 Tucanae. These contrasting pulsar populations may reflect the different evolutionary states and physical conditions of the two clusters. In particular, the central stellar density of Terzan 5 is about twice that of 47 Tucanae, resulting in many more stellar interactions that might disrupt the spin-up process in some binary systems and induce larger eccentricities in others.

Observations of the eccentric binaries I and J show that the point of closest approach of the two stars in each orbit changes with time. This relativistic effect is also seen in the orbit of Mercury, caused by the warping of spacetime close to the Sun. However, for I and J, the much stronger gravitational fields of the neutron stars mean that the rate of pre-

The author is at Jodrell Bank Observatory, University of Manchester, Macclesfield SK11 9DL, UK. E-mail: duncan.lorimer@manchester.ac.uk

Published by AAAS



Are more pulsars lurking in Terzan 5? (Left) A false-color optical image showing a 1 by 1 light-year field of the central region of Terzan 5 taken with the European Southern Observatory's New Technology Telescope. **(Right)** Radio map of the same region. [Adapted from (*8*)] Radio source Ter5A is the first pulsar found in this cluster (*5*). Most of the newly discovered pulsars are likely to contribute to the central continuum emission. The "unidentified radio source" might be a bright millisecond pulsar in a compact orbit.

cession in these binary systems is about 0.3° per year, some 2600 times that of Mercury. Interpreting this precession with general relativity provides strong constraints on the masses of the two pulsars and their companions. Ransom *et al.* (*1*) demonstrate that the pulsar masses are likely to be at least 1.7 times that of the Sun, at the high end of the distribution of measured neutron star masses (*9*). This result indicates that these neutron stars accreted large fractions of a solar mass during their formation. It also provides strong evidence for the existence of massive neutron stars, which are not allowed by some models of matter at high densities (*10*).

Terzan 5 is now a hive of activity for observational and theoretical astronomers. Within a year, further Green Bank observations will provide precise locations and spin-down rates of the pulsars within the cluster. These observations will allow comparative studies at optical and x-ray wavelengths and will probe the mass (*11*) and gas (*4*) distributions in the cluster. The 24 pulsars found so far only account for half of the radio emission seen in the second figure, and more are likely to be discovered with current instrumentation. Of particular interest is the nature of the radio source north of the cluster center (see the "unidentified radio source" in the second figure). This enigmatic source may be a bright millisecond pulsar that has so far evaded detection by virtue of being in an extremely compact orbit, perhaps around a black hole. The identification of such a system is a key goal of pulsar astronomy, and Terzan 5 may be the place to find it.

References

1. S. M. Ransom *et al.*, *Science* **307**, 892 (2005); published online 13 January 2005 (10.1126/science.1108632).
2. S. Sigurdsson, E. S. Phinney, *Astrophys. J. Suppl.* **99**, 609 (1995).
3. F. Camilo *et al.*, *Astrophys. J.* **535**, 975 (2000).
4. P. C. C. Freire *et al.*, *Astrophys. J.* **557**, L105 (2001).
5. A. G. Lyne *et al.*, *Nature* **347**, 650 (1990).
6. A. G. Lyne *et al.*, *Mon. Not. R. Astron. Soc.* **316**, 491 (2000).
7. S. M. Ransom, thesis, Harvard University (2001).
8. A. S. Fruchter, W. M. Goss, *Astrophys. J.* **536**, 865 (2000).
9. I. H. Stairs, *Science* **304**, 547 (2004).
10. J. M. Lattimer, M. Prakash, *Science* **304**, 536 (2004).
11. E. S. Phinney, *Philos. Trans. R. Soc. A* **341**, 39 (1992).

10.1126/science.1109555

MATHEMATICS/COMPUTATION

Accelerating Networks

John S. Mattick and Michael J. Gagen

Networks that are simple connection networks, such as telephone exchanges or the Internet, are able to grow in an unconstrained way. In contrast, regulatory networks—such as those in biology (for example, the network of regulatory proteins that controls gene expression in bacteria), engineering, or society—are accelerating networks that must be able to operate in a globally responsive way. Such global responsiveness, we argue, imposes an upper size limit on the complexity of integrated systems due to the costs incurred by the need for an increased number of connections and levels of regulation.

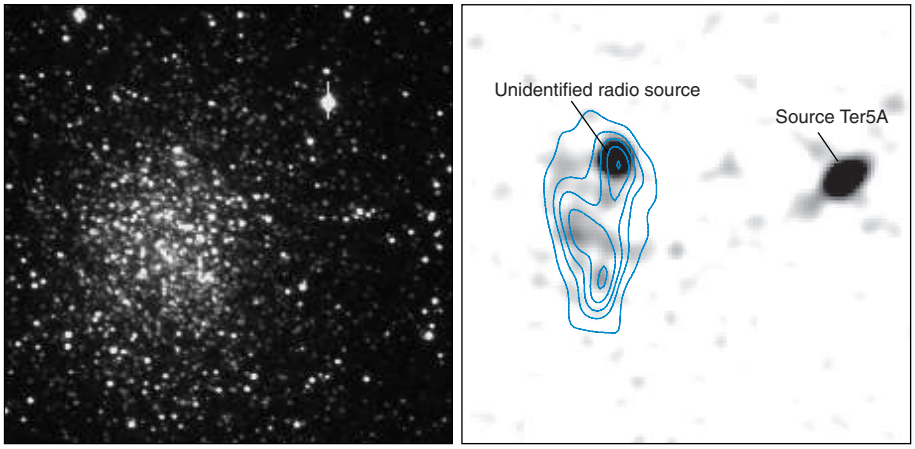
Most network studies to date have focused on simple (and usually large) connectionist systems like telephone exchanges or the Internet. These networks are generally "scale-free," that is, they exhibit little change in their network structures or statistics as they grow, in terms of the average number and the distribution of their connections per node (*1*, *2*). These networks can become large precisely because they have no need to rapidly integrate information from or globally respond to the current state of their nodes. For example, it does not matter to the overall function of the Internet whether any particular individual is connected or not. The state of one node is quite irrelevant to most of the others, although the system as a whole is vulnerable to damage at nodes that have many connections (*3*).

The situation is different with functionally organized systems whose operation is reliant on the integrated activity of any or all of its component nodes. Good examples include stock exchanges, the *Science* editorial office, and the protein network controlling gene expression. In such circumstances, the number of informative connections per node must increase with the size of the network. This means that the total number of connections between nodes scales faster than linearly with node number (see the figure, top). Such networks are termed "accelerating" networks (*1, 2*).

We contend that accelerating networks are far more common in the natural world than has hitherto been appreciated. These accelerating connection requirements, in principle and in practice, impose an upper limit on the functional complexity that integrated systems can attain. Maximal integrated connectivity occurs when all nodes are connected to all others (a proportional connectivity of 1), which means that the total number of connections in such networks scales quadratically with network size. Even if the proportional connectivity is much less than 1, the number of connections must still scale quadratically, otherwise global connectivity will decline. This in turn means that the size and complexity of such systems must sooner or later reach a limit where the number of possible connections becomes saturated or where the accelerating proportional cost of these connections becomes prohibitive.

This limit can be breached by a reduction in connectivity, which reduces the functional integration of the network, leading to fragmentation, as is observed, for example, in the

The authors are in the ARC Special Research Centre for Functional and Applied Genomics, Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland 4072, Australia. E-mail: j.mattick@imb.uq.edu.au, m.gagen@imb.uq.edu.au



Are more pulsars lurking in Terzan 5? (Left) A false-color optical image showing a 1 by 1 light-year field of the central region of Terzan 5 taken with the European Southern Observatory's New Technology Telescope. **(Right)** Radio map of the same region. [Adapted from (*8*)] Radio source Ter5A is the first pulsar found in this cluster (*5*). Most of the newly discovered pulsars are likely to contribute to the central continuum emission. The "unidentified radio source" might be a bright millisecond pulsar in a compact orbit.

cession in these binary systems is about 0.3° per year, some 2600 times that of Mercury. Interpreting this precession with general relativity provides strong constraints on the masses of the two pulsars and their companions. Ransom *et al.* (*1*) demonstrate that the pulsar masses are likely to be at least 1.7 times that of the Sun, at the high end of the distribution of measured neutron star masses (*9*). This result indicates that these neutron stars accreted large fractions of a solar mass during their formation. It also provides strong evidence for the existence of massive neutron stars, which are not allowed by some models of matter at high densities (*10*).

Terzan 5 is now a hive of activity for observational and theoretical astronomers. Within a year, further Green Bank observations will provide precise locations and spin-down rates of the pulsars within the cluster. These observations will allow comparative studies at optical and x-ray wavelengths and will probe the mass (*11*) and gas (*4*) distributions in the cluster. The 24 pulsars found so far only account for half of the radio emission seen in the second figure, and more are likely to be discovered with current instrumentation. Of particular interest is the nature of the radio source north of the cluster center (see the "unidentified radio source" in the second figure). This enigmatic source may be a bright millisecond pulsar that has so far evaded detection by virtue of being in an extremely compact orbit, perhaps around a black hole. The identification of such a system is a key goal of pulsar astronomy, and Terzan 5 may be the place to find it.

References

1. S. M. Ransom *et al.*, *Science* **307**, 892 (2005); published online 13 January 2005 (10.1126/science.1108632).
2. S. Sigurdsson, E. S. Phinney, *Astrophys. J. Suppl.* **99**, 609 (1995).
3. F. Camilo *et al.*, *Astrophys. J.* **535**, 975 (2000).
4. P. C. C. Freire *et al.*, *Astrophys. J.* **557**, L105 (2001).
5. A. G. Lyne *et al.*, *Nature* **347**, 650 (1990).
6. A. G. Lyne *et al.*, *Mon. Not. R. Astron. Soc.* **316**, 491 (2000).
7. S. M. Ransom, thesis, Harvard University (2001).
8. A. S. Fruchter, W. M. Goss, *Astrophys. J.* **536**, 865 (2000).
9. I. H. Stairs, *Science* **304**, 547 (2004).
10. J. M. Lattimer, M. Prakash, *Science* **304**, 536 (2004).
11. E. S. Phinney, *Philos. Trans. R. Soc. A* **341**, 39 (1992).

10.1126/science.1109555

MATHEMATICS/COMPUTATION

Accelerating Networks

John S. Mattick and Michael J. Gagen

Networks that are simple connection networks, such as telephone exchanges or the Internet, are able to grow in an unconstrained way. In contrast, regulatory networks—such as those in biology (for example, the network of regulatory proteins that controls gene expression in bacteria), engineering, or society—are accelerating networks that must be able to operate in a globally responsive way. Such global responsiveness, we argue, imposes an upper size limit on the complexity of integrated systems due to the costs incurred by the need for an increased number of connections and levels of regulation.

Most network studies to date have focused on simple (and usually large) connectionist systems like telephone exchanges or the Internet. These networks are generally "scale-free," that is, they exhibit little change in their network structures or statistics as they grow, in terms of the average number and the distribution of their connections per node (*1*, *2*). These networks can become large precisely because they have no need to rapidly integrate information from or globally respond to the current state of their nodes. For example, it does not matter to the overall function of the Internet whether any particular individual is connected or not. The state of one node is quite irrelevant to most of the others, although the system as a whole is vulnerable to damage at nodes that have many connections (*3*).

The situation is different with functionally organized systems whose operation is reliant on the integrated activity of any or all of its component nodes. Good examples include stock exchanges, the *Science* editorial office, and the protein network controlling gene expression. In such circumstances, the number of informative connections per node must increase with the size of the network. This means that the total number of connections between nodes scales faster than linearly with node number (see the figure, top). Such networks are termed "accelerating" networks (*1, 2*).

We contend that accelerating networks are far more common in the natural world than has hitherto been appreciated. These accelerating connection requirements, in principle and in practice, impose an upper limit on the functional complexity that integrated systems can attain. Maximal integrated connectivity occurs when all nodes are connected to all others (a proportional connectivity of 1), which means that the total number of connections in such networks scales quadratically with network size. Even if the proportional connectivity is much less than 1, the number of connections must still scale quadratically, otherwise global connectivity will decline. This in turn means that the size and complexity of such systems must sooner or later reach a limit where the number of possible connections becomes saturated or where the accelerating proportional cost of these connections becomes prohibitive.

This limit can be breached by a reduction in connectivity, which reduces the functional integration of the network, leading to fragmentation, as is observed, for example, in the

The authors are in the ARC Special Research Centre for Functional and Applied Genomics, Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland 4072, Australia. E-mail: j.mattick@imb.uq.edu.au, m.gagen@imb.uq.edu.au

transition of social networks from small communities to cities (*4*). However, if integration of node activity is absolutely required for the operation of the system or for its competitive survival, the functional complexity of the system can only be increased beyond the existing limit by increasing the number of connections. This can be accomplished by changing the physical basis of the connections and reducing their proportionate cost (see the figure, bottom).

Computer systems are the preeminent example of accelerating networks. The millions of components on an integrated chip must be connected together via common bus lines requiring hundreds of meters of metal wires to be layered across a thumbnail-sized chip. These burdensome connectivity requirements limit chip and motherboard component densities, speeds, and layouts, all of which impede performance (*5*). To get around such problems, typical supercomputers may be configured as clusters of smaller computers (*6*) or a virtual grid (*7*), but the size and complexity of the implemented supercomputer are still limited by its connectivity. Clusters and grid computers must balance overall communication bandwidth constraints, latency requirements, and the need for parallel load balancing, all of which combine to confer constraints on scalability. The tighter these constraints and the more powerful the supercomputer, the greater the proportion of the structure and cost that must be devoted to addressing these connection issues. The nonlinear growth of the additional costs of regulation and physical connectivity limits conspire to constrain the size of supercomputers.

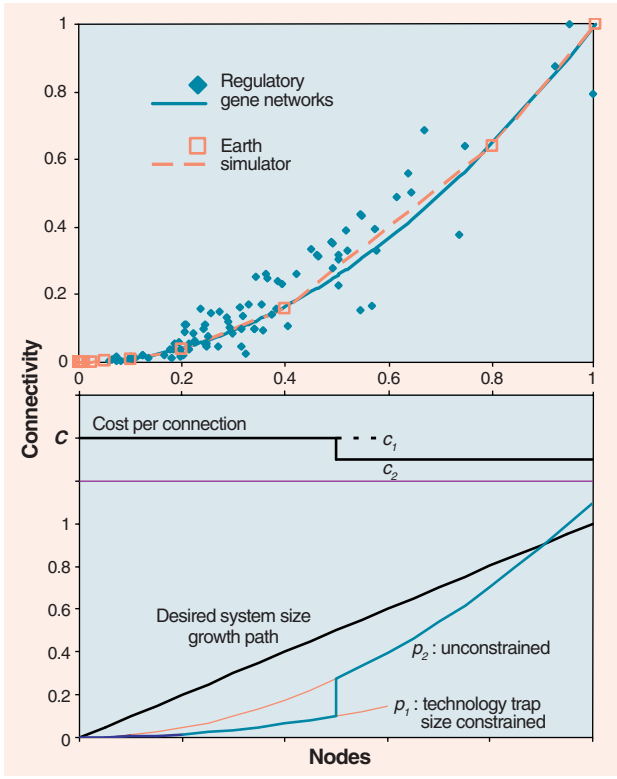
Similar limits occur in industrial systems such as integrated production systems, and in biology. For example, the Toyota Just-in-Time production system, now used widely by many industries, relies not only on an integrated assembly line but also on the real-time delivery of externally manufactured subcomponents (*8*), which increases cost efficiency but also increases the fragility of the system. In stock exchanges, communication was, until relatively recently, based upon an open outcry system whereby traders on the floors would yell to others. This method of connection ultimately restricts the number of specialist traders, the information available to them, and the size and efficiency of the trading community. The need to surmount the inherent limits of an open outcry system has forced most major exchanges to automate their trading operations in the interests of obtaining a larger and more functional (better connected) exchange.

When connection limits cannot be raised, or functional components cannot directly communicate with each other, the alternative is to introduce dedicated hierarchies—called management in organizations, control systems in engineering, and regulation in biology. These regulatory systems still scale nonlinearly (albeit not quadratically) with system size in a manner that is dependent on the number of nodes being managed at each step. Moreover, these hierarchies have their own costs. Each level of regulatory hierarchy introduces time delays, increases noise and stochastic errors, and results in loss of detailed information from within subnetworks across the system. These shortcomings increase with greater levels of regulation and with network size (that is, the bigger the organization, the greater the number of management levels), limiting system coherence and ultimately imposing upper limits on the size and functional complexity that such systems can attain. Evidently, the most effective network organization is the one that is the most completely interconnected.

Living organisms are highly integrated systems. The same logic may explain the empirical observation that the number of regulatory proteins controlling gene expression in bacteria increases quadratically with genome size, and may be the cause of the observed upper limit on the size of bacterial genomes (*9, 10*). In any highly competitive system—whether biological or industrial—the speed and efficiency of organization, and the sophistication of response to changing circumstances are critical determinants of the system's survival and success. We suggest that this is the imperative that results in biological regulatory networks scaling quadratically with system size in order to maintain optimal integration.

Systems that require integral organization to function in a competitive environment are dependent on, and ultimately constrained by, their accelerating regulatory architecture. Thus, connectivity and the proportion of the system devoted to regulation must scale faster than function in organized complex systems. Moreover, any network that requires time-critical integration to operate and to be competitive suffers decreasing benefits as size increases. The decrease in benefits can be in terms of connection costs or organizational costs (information loss, increases in noise and processing time), which ultimately constrain the size and complexity of the network.

These constraints can only be relieved by technological transitions that alter the physical basis of the control architecture. Thus, accelerating networks show quasi-stationary phases of growth in their complexity and capability, asymptotically approaching maxima until the ceiling is lifted (see the figure, bottom). Such growth may itself involve the development of some new technology whose conception and implementation are constrained by other parameters, usually knowledge.



The costs and limits of growth. (**Top**) Both supercomputers [open squares, dashed line (*14, 15*)] and single-celled prokaryotic organisms [blue diamonds, solid line (*9, 10*)] increase their network size (horizontal axis) through quadratic increases in the number of network connections (connectivity; vertical axis), as depicted in this scaled plot. For example, within a network of 4000 bacterial genes, ~2.5% of the genes are regulatory, whereas in a genome of 8000 genes, 9% of the genes are regulatory. Meanwhile, Japan's Earth Simulator, one of the world's fastest supercomputers, requires more than 400,000 connections carried by 83,000 wires to connect 640 nodes. (**Bottom**) Whether the system is the "Earth Simulator," a bacterium, or a business organization, the costs per connection determine the point at which the decreasing returns with increase in size end up limiting the ultimate size of the system. These inherent limits can only be lifted by introducing new technologies that significantly decrease connection costs.

Published by AAAS

Consider, for instance, the proliferation of digital programming in such domains as computing and aviation. The increased sophistication of modern aircraft has been enabled not so much by advances in their analog components, but rather by transitions from mechanical to computational control systems, including software and optical fibers, which now constitute much of the cost (*11*). Indeed, the transition to digital technologies is a generic solution to previous limitations of analog communication and control technologies. In virtually all systems, it has been observed that explosive increases in complexity occur as a result of more advanced controls and embedded networking, most of which is invisible to the observer (*11*). In the case of biology, we suggest that the transition from an analog protein-based regulatory system to a digital RNA-based control architecture was the critical platform that enabled the evolution and development of complex organisms (*12, 13*).

There are very few, if any, fully scalable technologies. Software programming that suffices for a small program of a few hundred lines is unable to produce workable programs of millions of lines. Management structures that suffice for small organizations do not readily scale to large ones. Biological organisms are a collection of technologies optimal in some respects but suboptimal in others, which limits life's potential. Understanding where the points of regulatory saturation and technological limitation occur will be necessary to break through present and future barriers to increases in complexity.

References

1. S. N. Dorogovtsev, J. F. F. Mendes, *Adv. Phys.* **51**, 1079 (2002).
2. R. Albert, A. L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
3. R. Albert, H. Jeong, A. L. Barabási, *Nature* **406**, 378 (2000).
4. M. E. J. Newman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2566 (2002).
5. R. Sommerhalder, S. C. van Westrhenen, *The Theory of Computability: Programs, machines, effectiveness and feasibility* (Addison-Wesley, Workingham, UK, 1988).
6. W. W. Hargrove *et al.*, *Sci. Am.* **285** (2), 73 (2001).
7. I. Foster. *Sci. Am.* **288** (4), 79 (2003).
8. J. P. Womack *et al. The Machine that Changed the World* (Rawson Associates, New York, 1990).
9. L. J. Croft *et al.*, http://arxiv.org/abs/q-bio.MN/0311021 (2003).
10. M. J. Gagen, J. S. Mattick, *Theory Bioscienc.*, in press.
11. M. E. Csete, J. C. Doyle, *Science* **295**, 1664 (2002).
12. J. S. Mattick, M. J. Gagen, *Mol. Biol. Evol.* **18**, 1611 (2001).
13. J. S. Mattick, *Nat. Rev. Genet.* **5**, 316 (2004).
14. D. J. Kerbyson, A. Hoisie, H. Wasserman, Proceedings of the International Parallel and Distributed Processing Symposium (IPDPS'03), Nice, France, 22 to 26 April 2003.
15. S. Habata *et al.*, *NEC Res. Dev.* **44**, 21 (2003).

10.1126/science.1103737

IMMUNOLOGY

Thymic Regulation—Hidden in Plain Sight

Ellen V. Rothenberg

The thymus gland in the neck is the nursery where precursor T cells differentiate into fully fledged mature T lymphocytes. A long-standing mystery about the thymus concerns why it continuously overproduces immature T cells of which only a tiny fraction are permitted to survive to maturity. Most of these immature lymphocytes—called double positive (DP) cells because they express both CD4 and CD8—begin to express αβ T cell receptors (TCRs) in the thymus but are remarkably inert when tested in most assays that measure functional T cell responses. DP cells, each expressing a different αβ TCR, compete for the rare chance to survive based on the usefulness of their individual TCRs. Traditionally, they have been viewed as auditioning for the T cell recognition repertoire, being overproduced and then killed off only to provide maximum potential TCR diversity. However, as Silva-Santos *et al*. reveal on page 925 of this issue (*1*), DP immature lymphocytes also are functionally active in another capacity. They organize and modulate thymic function by secreting factors that guide maturation of other T cell lineages and feed back on the gene-expression programs of the next generation of T cell progenitors. As Silva-Santos *et al*. demonstrate, they do this by using some of the same molecular mechanisms that organize the development of peripheral lymph nodes in the embryo (*1*).

Enhanced online at www.sciencemag.org/cgi/content/full/307/5711/858

The author is in the Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA. E-mail: evroth@its.caltech.edu

The new findings of Silva-Santos *et al*. build on the group's previous work, which defined molecular differences between two major lineages of T cells: the TCRαβ cells and the TCRγδ cells (*2*). Although the TCRαβ cells are more familiar and abundant among circulating lymphocytes, the TCRγδ cells are crucial for host defense and wound healing at epithelial surfaces. The development of TCRγδ cells branches off from that of TCRαβ cells during the double negative (DN) stages of T cell development in the thymus (when the T cell progenitors do not express CD4 or CD8). The authors identified a suite of genes preferentially expressed in TCRγδ cells, but with expression entirely dependent on the context in which the cells developed. Expression of these genes was switched on in TCRγδ cells only if TCRαβ DP cells were also developing in the same thymus (*2*). In their new work, these investigators reveal that the correlation is causal: By secreting lymphotoxin, the DP cells induce their neighbors to express this special set of γδ-associated genes. Furthermore, the ability to make this lymphotoxin efficiently depends on a transcription factor, RORγt, which is only expressed by thymocytes in the DP population (*1*).

This work sheds new light on DP cell programming. DP cells cannot express classic cytokine genes or effector genes (such as those encoding perforin or granzymes), and only the few that are selected for maturation seem to acquire these critical capabilities. The incompetence of DP cells in all aspects of mature T cell function is often regarded as a sign of their immaturity, but in fact they are specifically disabled. The ability to produce cytokines and other functions of DN precursors are lost after they first express TCRβ chains and undergo the transition from DN to DP lymphocytes (*3*). This transition, known as β selection, occurs only in TCRαβ lineage precursors and separates them definitively from the TCRγδ cell lineage. This transition is characterized by a burst in proliferation and a cascade of regulatory changes.

One transcription factor, whose expression is turned on during β selection, is RORγt. This factor is important for helping to direct the distinctive physiology of TCRαβ DP cells in that it provides a DP-specific survival mechanism (*4*). RORγt also acts as a powerful temporary repressor of mature T cell effector function, and must be turned off again to restore this function in the rare DP cells that are selected to mature (*5*). Silva-Santos *et al*. now show that RORγt is also a positive regulator of one function, directly or indirectly enabling DP cells to act on their neighbors by secreting lymphotoxin. Thus, RORγt may be more of a switch between alternative sets of functions than a simple inhibitor.

Consider, for instance, the proliferation of digital programming in such domains as computing and aviation. The increased sophistication of modern aircraft has been enabled not so much by advances in their analog components, but rather by transitions from mechanical to computational control systems, including software and optical fibers, which now constitute much of the cost (*11*). Indeed, the transition to digital technologies is a generic solution to previous limitations of analog communication and control technologies. In virtually all systems, it has been observed that explosive increases in complexity occur as a result of more advanced controls and embedded networking, most of which is invisible to the observer (*11*). In the case of biology, we suggest that the transition from an analog protein-based regulatory system to a digital RNA-based control architecture was the critical platform that enabled the evolution and development of complex organisms (*12, 13*).

There are very few, if any, fully scalable technologies. Software programming that suffices for a small program of a few hundred lines is unable to produce workable programs of millions of lines. Management structures that suffice for small organizations do not readily scale to large ones. Biological organisms are a collection of technologies optimal in some respects but suboptimal in others, which limits life's potential. Understanding where the points of regulatory saturation and technological limitation occur will be necessary to break through present and future barriers to increases in complexity.

References

1. S. N. Dorogovtsev, J. F. F. Mendes, *Adv. Phys.* **51**, 1079 (2002).
2. R. Albert, A. L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
3. R. Albert, H. Jeong, A. L. Barabási, *Nature* **406**, 378 (2000).
4. M. E. J. Newman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2566 (2002).
5. R. Sommerhalder, S. C. van Westrhenen, *The Theory of Computability: Programs, machines, effectiveness and feasibility* (Addison-Wesley, Workingham, UK, 1988).
6. W. W. Hargrove *et al.*, *Sci. Am.* **285** (2), 73 (2001).
7. I. Foster. *Sci. Am.* **288** (4), 79 (2003).
8. J. P. Womack *et al. The Machine that Changed the World* (Rawson Associates, New York, 1990).
9. L. J. Croft *et al.*, http://arxiv.org/abs/q-bio.MN/0311021 (2003).
10. M. J. Gagen, J. S. Mattick, *Theory Bioscienc.*, in press.
11. M. E. Csete, J. C. Doyle, *Science* **295**, 1664 (2002).
12. J. S. Mattick, M. J. Gagen, *Mol. Biol. Evol.* **18**, 1611 (2001).
13. J. S. Mattick, *Nat. Rev. Genet.* **5**, 316 (2004).
14. D. J. Kerbyson, A. Hoisie, H. Wasserman, Proceedings of the International Parallel and Distributed Processing Symposium (IPDPS'03), Nice, France, 22 to 26 April 2003.
15. S. Habata *et al.*, *NEC Res. Dev.* **44**, 21 (2003).

10.1126/science.1103737

IMMUNOLOGY

Thymic Regulation—Hidden in Plain Sight

Ellen V. Rothenberg

The thymus gland in the neck is the nursery where precursor T cells differentiate into fully fledged mature T lymphocytes. A long-standing mystery about the thymus concerns why it continuously overproduces immature T cells of which only a tiny fraction are permitted to survive to maturity. Most of these immature lymphocytes—called double positive (DP) cells because they express both CD4 and CD8—begin to express αβ T cell receptors (TCRs) in the thymus but are remarkably inert when tested in most assays that measure functional T cell responses. DP cells, each expressing a different αβ TCR, compete for the rare chance to survive based on the usefulness of their individual TCRs. Traditionally, they have been viewed as auditioning for the T cell recognition repertoire, being overproduced and then killed off only to provide maximum potential TCR diversity. However, as Silva-Santos *et al*. reveal on page 925 of this issue (*1*), DP immature lymphocytes also are functionally active in another capacity. They organize and modulate thymic function by secreting factors that guide maturation of other T cell lineages and feed back on the gene-expression programs of the next generation of T cell progenitors. As Silva-Santos *et al*. demonstrate, they do this by using some of the same molecular mechanisms that organize the development of peripheral lymph nodes in the embryo (*1*).

Enhanced online at www.sciencemag.org/cgi/content/full/307/5711/858

The author is in the Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA. E-mail: evroth@its.caltech.edu

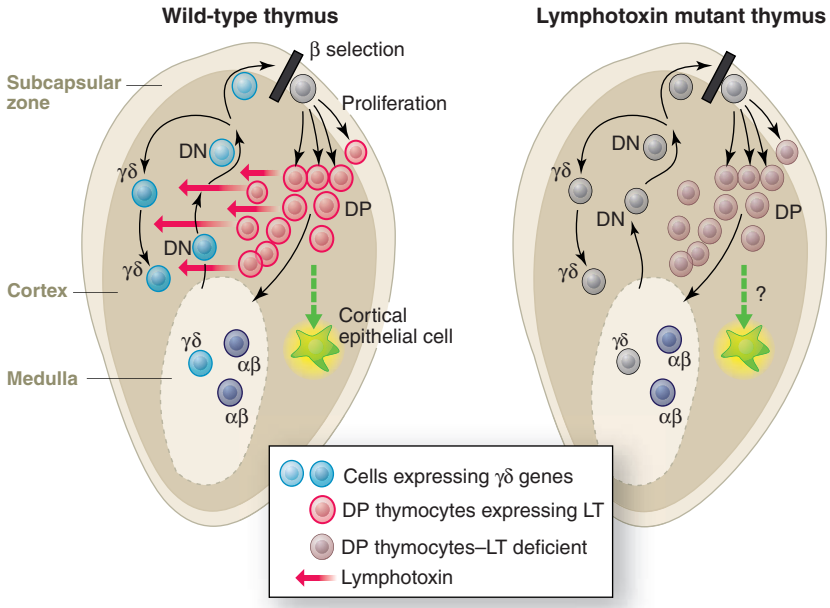
The new findings of Silva-Santos *et al*. build on the group's previous work, which defined molecular differences between two major lineages of T cells: the TCRαβ cells and the TCRγδ cells (*2*). Although the TCRαβ cells are more familiar and abundant among circulating lymphocytes, the TCRγδ cells are crucial for host defense and wound healing at epithelial surfaces. The development of TCRγδ cells branches off from that of TCRαβ cells during the double negative (DN) stages of T cell development in the thymus (when the T cell progenitors do not express CD4 or CD8). The authors identified a suite of genes preferentially expressed in TCRγδ cells, but with expression entirely dependent on the context in which the cells developed. Expression of these genes was switched on in TCRγδ cells only if TCRαβ DP cells were also developing in the same thymus (*2*). In their new work, these investigators reveal that the correlation is causal: By secreting lymphotoxin, the DP cells induce their neighbors to express this special set of γδ-associated genes. Furthermore, the ability to make this lymphotoxin efficiently depends on a transcription factor, RORγt, which is only expressed by thymocytes in the DP population (*1*).

This work sheds new light on DP cell programming. DP cells cannot express classic cytokine genes or effector genes (such as those encoding perforin or granzymes), and only the few that are selected for maturation seem to acquire these critical capabilities. The incompetence of DP cells in all aspects of mature T cell function is often regarded as a sign of their immaturity, but in fact they are specifically disabled. The ability to produce cytokines and other functions of DN precursors are lost after they first express TCRβ chains and undergo the transition from DN to DP lymphocytes (*3*). This transition, known as β selection, occurs only in TCRαβ lineage precursors and separates them definitively from the TCRγδ cell lineage. This transition is characterized by a burst in proliferation and a cascade of regulatory changes.

One transcription factor, whose expression is turned on during β selection, is RORγt. This factor is important for helping to direct the distinctive physiology of TCRαβ DP cells in that it provides a DP-specific survival mechanism (*4*). RORγt also acts as a powerful temporary repressor of mature T cell effector function, and must be turned off again to restore this function in the rare DP cells that are selected to mature (*5*). Silva-Santos *et al*. now show that RORγt is also a positive regulator of one function, directly or indirectly enabling DP cells to act on their neighbors by secreting lymphotoxin. Thus, RORγt may be more of a switch between alternative sets of functions than a simple inhibitor.

The responses to the lymphotoxin-mediated pathway demonstrated here (*1*) are not confined to TCRγδ cells. One of the most interesting effects of this pathway is on early DN cells, that is, cells that probably have not yet reached the branch point between the TCRγδ and TCRαβ fates. These early cells also express the "γδ-associated" genes if DP cells are present, but not if DP cells are absent (*1*). The picture of intrathymic traffic patterns revealed in the past few years by Petrie and co-workers (*6*) helps to illuminate what this means. T cell precursors in the DN stages migrate outward through the cortex of the thymus as they begin to rearrange their TCR genes (see the figure). They undergo β selection at the outermost periphery of the thymic cortex, and then, as DP cells, they fall back down through the cortex toward the medulla. The most immature precursors must therefore navigate their way among a crush of DP cells going in the opposite direction. This scheme suggests obvious possibilities for homeostatic regulation. Indeed, there is evidence that the flux of precursors through the earliest stages in the thymus might be controlled by some kind of feedback mechanism (*7*). Until now, there was little if any molecular evidence that DN cells altered the expression of specific genes in response to the presence of DP cells. Silva-Santos *et al*. set an important precedent by demonstrating the effects of DP cell–generated lymphotoxin on the gene-expression patterns of these early precursors.

These results also illuminate new aspects of the divergence between TCRαβ and TCRγδ lineages. The "γδ-associated" genes are not really specific to TCRγδ cells, but rather may be an inheritance from earlier DN precursors. Instead, it is the TCRαβ cells, as they enter the DP stage, that break from the precursor gene-expression pattern. They presumably shut off any "γδ-associated" gene expression they may have inherited from their own precursors; but then they turn on expression of the genes encoding RORγt and lymphotoxin in order to induce "γδ-associated" genes in neighboring cells that will be the next generation of precursors. A current model for TCRγδ and TCRαβ divergence suggests that strength of signals through the nascent TCR complex determines lineage choice, with stronger signals leading to a TCRγδ fate and weaker ones leading to a TCRαβ fate (*8*). Even in the absence of TCRβ, some cells make the DN-to-DP transition, probably through transient expression of TCRγδ receptors that signal weakly. However, this abnormal β selection does not equip the cells for efficient expression of either RORγt or lymphotoxin genes (*1*). Thus, a DP phenotype, although necessary, is not sufficient for activation of the lymphotoxin pathway.



A thymic Virginia Reel. Lymphotoxin (LT) secreted by DP thymocytes after β selection triggers expression of the γδ-associated set of genes in both γδ thymocytes and DN precursor thymocytes. (**Left**) Pathways of migration through the thymus during the DN and DP stages of differentiation. In the normal thymus, DP cells activate the γδ-associated gene set in DN precursors and their TCRγδ descendants through a lymphotoxin-dependent pathway (red arrows). DP cells also influence epithelial cells of the thymic cortex (green arrows). (**Right**) In the absence of lymphotoxin, expression of γδ-associated genes is not turned on even though TCRγδ cells are generated in normal numbers. Defective induction also occurs in an RORγt mutant thymus. Atypical DP cells generated by abnormal β selection also fail to produce adequate amounts of RORγt and lymphotoxin.

Through their production of RORγt and lymphotoxin, DP cells are linked to a different cell lineage, that of lymphoid tissue inducer (LTi) cells. These cells have a powerful role in the immune system as they direct the formation of peripheral lymph nodes in the developing fetus (*9*). They develop along with lymphocytes from a common fetal precursor cell (*10*), but their developmental program is thought to diverge from B and T cells at an early stage because they depend on the inhibitory transcription factor Id2, which blocks T and B cell development (*11*). LTi cells also express RORγt and use lymphotoxin as their main intercellular communicator to induce and organize differentiation of lymph nodes from mesenchymal tissue. Strikingly, after extended T cell lineage differentiation, DP cells still have access to this "atavistic" effector pathway.

The sharing of these genetic circuit elements between LTi and DP cells may be one molecular explanation for years of data indicating that DP cells also possess some mechanism to organize the morphogenesis of stroma in their intrathymic environment (*12*). A recent report showed that the presence of DP cells may determine where in the thymic cortex epithelial cells will be allowed to secrete the cytokine interleukin-7 (IL-7) (*13*). IL-7 is a key proliferative regulator for developing T cells, and also is required for the generation of TCRγδ cells. It could be important for feedback by DP cells on the initial entry of T cell precursors into the thymic cortex, as well as on TCRγδ maturation. It remains to be seen whether DP cells use the same lymphotoxin-mediated pathway to control IL-7 secretion by their epithelial neighbors as they now seem to use to modulate their lymphoid neighbors. In either case, it is a long way from idle passivity in the waiting room of the TCR audition hall.

References

1. B. Silva-Santos, D. J. Pennington, A. C. Hayday, *Science* **307**, 925 (2005); published online 9 December 2004 (10.1126/science.1103978).
2. D. J. Pennington *et al.*, *Nature Immunol.* **4**, 991 (2003).
3. E. V. Rothenberg, M. A. Yui, J. C. Telfer, in *Fundamental Immunology 5th edition*, W. E. Paul, Ed. (Lippincott, Williams & Wilkins, Philadelphia, 2003), chap. 9.
4. Z. Sun *et al.*, *Science* **288**, 2369 (2000).
5. Y.-W. He *et al.*, *J. Immunol.* **164**, 5668 (2000).
6. H. T. Petrie, *Nature Rev. Immunol.* **3**, 859 (2003).
7. S. E. Prockop, H. T. Petrie, *J. Immunol.* **173**, 1604 (2004).
8. S. M. Hayes, E. W. Shores, P. E. Love, *Immunol. Rev.* **191**, 28 (2003).
9. G. Eberl *et al.*, *Nature Immunol.* **5**, 64 (2004).
10. R. E. Mebius *et al.*, *J. Immunol.* **166**, 6593 (2001).
11. Y. Yokota *et al.*, *Nature* **397**, 702 (1999).
12. P. Naquet *et al.*, *Semin. Immunol.* **11**, 47 (1999).
13. M. Zamisch *et al.*, *J. Immunol.* **174**, 60 (2005).

10.1126/science.1109501

Toward Atom Chips

József Fortágh and Claus Zimmermann

The spectacular success of microelectronics has demonstrated the enormous potential of miniaturization for turning basic physics into applications. Today, researchers are exploring further miniaturization to nanometer and even atomic scales. A new field of research explores the behavior of clouds of ultracold atoms (which behave as matter waves) above the surface of a microchip. The chip generates a magnetic and electrostatic field that can be used to guide and manipulate the matter wave. This setup may allow the construction of matter wave interferometers on microchips; such "atom chips" may serve as sensitive probes for gravity, acceleration, rotation, and tiny magnetic forces.

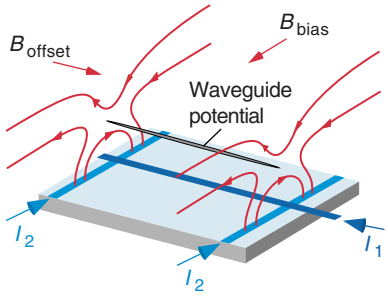
In 1995, Weinstein and Libbrecht published a simple and striking idea (*1*) for the storage, guidance, and manipulation of ultracold atoms near the surface of a chip containing microfabricated conductors. At the time, thermal clouds of ultracold atoms with temperatures below 0.001 K were routinely stored in magnetic traps. It was conceivable that the required magnetic storage fields could also be generated with microfabricated wires. But the real thrill was the vision of a new kind of chip based on atoms that behave like quantum mechanical matter waves rather than classical particles. However, atomic clouds could not yet be prepared at the very low temperatures required for quantized motion.

The discovery of Bose-Einstein condensation provided a possible solution. In such a condensate, the matter waves of the atoms overlap and all the atoms are indistinguishable, forming a macroscopic matter wave. A microtrap would be able to carry a condensate of several tens of thousands of atoms. In 2001, two groups succeeded independently in loading Bose-Einstein condensates into a microtrap (*2*, *3*). The door was opened for the field of integrated atom optics (*4*–*6*).

Magnetic trapping requires the magnetic moment of the atoms to be antiparallel to the local magnetic field. The atoms then become trapped at the local minimum of the absolute value of the field. Different microtraps can be constructed. In the simplest microtrap, the circular magnetic field of a straight conductor is added to a homogeneous bias field oriented perpendicular to the conductor (see the first figure). The two fields compensate each other along a line that runs parallel to the conductor and defines a "waveguide potential." The atoms are trapped along this line.

In typical microtrap experiments, the conductor is driven with a few tens of milliamps, and the bias field is a few gauss. Under these conditions, the waveguide potential is separated from the conductor surface by several micrometers. The atom cloud thus floats above the chip and is thermally isolated from it, because the experiments are performed in a vacuum.

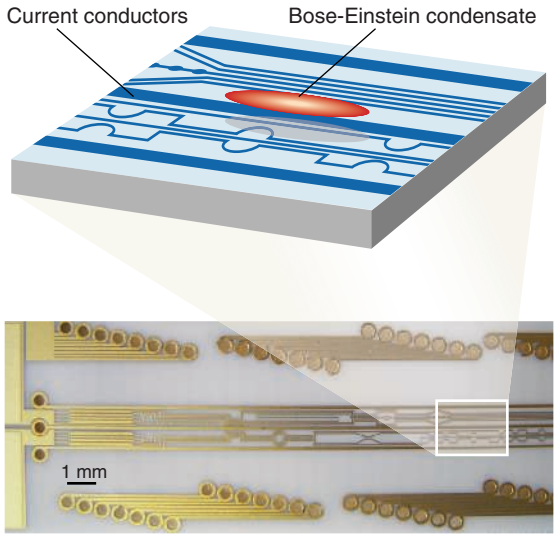
The miniaturization of the traps results in extremely strong magnetic field gradients and steep potentials. When a magnetic offset field is added parallel to the waveguide, the radial confinement becomes parabolic, giving rise to harmonic oscillations (see the first figure). For rubidium, the most commonly used element on atom chips, oscillation frequencies of 1000 Hz to tens of thousands of Hz are accessible. The potential structure along the axis of the waveguide can be manipulated by adding further conductors perpendicular to the waveguide axis. For example, a pair of additional conductors would close the waveguide at each end, resulting in three-dimensional confinement on a chip. More complex structures can be used to control the momentum and shape of the atomic matter waves (see the second figure) (*7*).



How to create a magnetic waveguide. The current-carrying conductor (I_1, dark blue) generates a circular magnetic field. When a homogeneous bias field B_{bias} is added, a line of vanishing magnetic field results. This line is the waveguide potential along which the atoms in an atom chip are trapped. A further offset field B_{offset} parallel to the conductor causes the radial confinement to become harmonic. The magnetic field of the perpendicular conductors (I_2, bright blue) closes the ends of the waveguide.

The steepness of the magnetic trapping potential determines the separation of the energy levels in the trap and sets the time scale for the atomic dynamics. Fast "circuits" require tiny traps, and tightly confining traps made of very thin conductors are thus desirable.

But how small can the traps be made? The fabrication of conductors down to the micrometer scale causes no particular difficulty. However, interactions of the conductors and the substrate surface (which are at room temperature) with the ultracold atoms must be considered. Theoretical considerations suggest that when the distance between the chip and the atom cloud is less than 1 μm, the lifetime of the trap should be reduced substantially by magnetic field fluctuations (*8*). Surprisingly, pronounced surface effects already occur at much larger distances of 100 μm and more. If a condensate is released into a waveguide potential, it does not form a smoothly expanding atomic cloud but fragments into little localized blobs. This effect (*9*) is probably caused by geometric imperfections and inhomogeneities in the conductor that force the current to deviate from a straight line. As a result, the waveguide becomes rough and bumpy. Condensates may thus serve as very sensitive probes for



Bose-Einstein condensate on a microchip. (Top) The wider conductors (dark blue) generate a waveguide for the atomic cloud (a Bose-Einstein condensate, shown enlarged by 10 times). The finer conductors (bright blue) generate micropotentials such as beam splitters, double-well potentials, and lattices. They can be loaded by adjusting the current in the outer conductors (*7*). **(Bottom)** Photograph of the complete atom chip (*7*), a section of which is shown schematically in the top panel.

The authors are with the Physikalisches Institut, Eberhard Karls Universität Tübingen, 72076 Tübingen, Germany. E-mail: fortagh@pit.physik.uni-tuebingen.de, clz@pit.physik.uni-tuebingen.de

Published by AAAS

sensing the path of a current in a conductor or in microfabricated heterostructures.

Another limit to miniaturization is set by the electrostatic attraction between atoms that are adsorbed and polarized at the chip surface. These forces become effective at micrometer-scale distances. A final limit is set by weak, attractive dispersion forces between the chip surface and the atom cloud. They rapidly set in at distances of several hundred nanometers and are difficult to compensate because of their strong spatial dependence.

Given these limits, there are several directions for future developments. Microfabrication can be improved to reduce imperfections and obtain more homogeneous current conductors—for example, by switching from electroplating to direct vapor deposition. However, basic atom chips can already be created with existing technology. Spatial fluctuations of the magnetic field due to imperfections of the conductors average out over distances that are large relative to the size of the imperfections. A waveguide far from the chip surface exhibits a smooth potential. This smooth waveguide can be combined with steep and thin potential barriers, generated on an extra chip next to the atomic cloud. Such a geometry, in which waveguide and potential barriers are generated by separate chips, is far less sensitive to wire imperfections than a single-layer chip. Matter-wave interferences have recently been observed with such a setup (*10*).

Atomic matter waves on chips can also be manipulated with electric fields (*11*) or even laser beams (*12*). Both induce an electric dipole moment through which the interaction with atoms is achieved. Combinations of magnetic, electrostatic, and optical potentials on atom chips remain to be explored. For example, temporally controlled lasers can be used to produce potential barriers in magnetic waveguides.

Instead of manipulating a condensate as a whole, it can also be used as a zero-temperature reservoir for single atoms. Such an atom laser could provide a coherent source for performing optics with single atoms. To measure and detect single atoms after they have been manipulated, a detector with single-atom sensitivity would be desirable. Possible approaches include optical ionization with light from the tip of an optical fiber followed by ion detection, and detection of single atoms in miniaturized optical resonators. According to theoretical studies, present-day technology should permit the construction of microcavities that can detect a single atom in the trap (*13*).

Which of the various ideas will prove most successful remains to be shown. As in every young field of research, atom chips will be good for future surprises.

References

1. J. D. Weinstein, K. G. Libbrecht, *Phys. Rev. A* **52**, 4004 (1995).
2. H. Ott, J. Fortágh, G. Schlotterbeck, A. Grossmann, C. Zimmermann, *Phys. Rev. Lett.* **87**, 230401 (2001).
3. W. Hänsel, P. Hommelhoff, T. W. Hänsch, J. Reichel, *Nature* **413**, 498 (2001).
4. W. Hänsel *et al.*, *Phys. Rev. A* **64**, 063607 (2001).
5. A. Smerzi *et al.*, *Phys. Rev. Lett.* **79**, 4950 (1997).
6. U. Dorner *et al.*, *Phys. Rev. Lett.* **91**, 073601 (2003).
7. J. Fortágh *et al.*, *Opt. Commun.* **243**, 45 (2004).
8. C. Henkel, S. Pötting, M. Wilkens, *Appl. Phys. B* **69**, 379 (1999).
9. J. Estève *et al.*, *Phys. Rev. A* **70**, 043629 (2004).
10. J. Fortágh, Institute for Theoretical Atomic, Molecular and Optical Physics (ITAMP) Workshop on Quantum Degenerate Gases in Low-Dimensionality, Cambridge, MA, 4 to 6 October 2004 (http://itamp.harvard.edu/low-d/low-d.html).
11. P. Krüger *et al.*, *Phys. Rev. Lett.* **91**, 233201 (2003).
12. Y.-J. Wang *et al.*, http://arxiv.org/abs/cond-mat/0407689.
13. P. Horak *et al.*, *Phys. Rev. A* **67**, 043806 (2003).

10.1126/science.1107348

PALEONTOLOGY

Homoplasy in the Mammalian Ear

Thomas Martin and Zhe-Xi Luo

The similarity among structures that arose through independent evolution instead of descent from a common ancestor is termed homoplasy and is a major feature of evolutionary morphology. A fascinating but very difficult question facing evolutionary biologists is whether a complex structure would be less likely than a simple structure to undergo independent homoplastic evolution (*1*). On page 910 of this issue, Rich *et al*. (*2*) partially answer this question with their analysis of the dentary bone from the lower jaw of an Early Cretaceous fossil monotreme called *Teinolophos*, an extinct relative of Australia's modern platypus and echidna. The new fossil find offers fresh anatomical evidence to support the hypothesis that a key evolutionary innovation among modern mammals—the separation of the middle ear bones from the mandible—must have evolved independently among the monotreme mammals and the therians (marsupials and placentals).

T. Martin is at the Forschungsinstitut Senckenberg, D-60325 Frankfurt am Main, Germany. E-mail: tmartin@senckenberg.de Z.-X. Luo is at the Carnegie Museum of Natural History, Pittsburgh, PA 15213, USA. E-mail: luoz@carnegiemnh.org

The tiny bones of the middle ear that are used for hearing render modern mammals—including placentals, pouched marsupials, and egg-laying monotremes—unique among vertebrates (*2*). The middle ear bones are the malleus, incus, and stapes, and in addition there is the tympanic bone, which supports the tympanic membrane, enabling it to receive sound. The tympanic, malleus, and incus are homologous to bones in the mandible and jaw hinge (the angular, articular, and quadrate, respectively), which are required for feeding in nonmammalian vertebrates (*3*–*5*). There is also extensive evidence from fossils of extinct cynodont and mammaliaform relatives of modern mammals suggesting that the angular, articular, and quadrate bones in these creatures were used for hearing while still attached to the mandible and jaw hinge (*6*, *7*). Evolution of the mammalian jaw joint and middle ear represents a classic example of the phylogenetic transformation of a complex functional structure that can be read directly from fossil evidence.

However, alternative interpretations have waxed and waned about how these middle ear bones got separated from the mandible during early mammalian evolution. The structure of the middle ear is so complex and unique that some researchers consider the separation of the middle ear bones from the mandible to be the strongest synapomorphic (shared derived) characteristic of living mammals (*7*, *8*). They also propose that the monotreme, marsupial, and placental lineages split after their common ancestor had acquired this key feature. This view has been contested by others who favor multiple and independent acquisitions of the mammalian middle ear bones after the divergence of monotremes, marsupials, and placentals (*6*, *9*).

In premammalian cynodonts and such primitive mammaliaforms as *Morganucodon* (see the figure), the middle ear bones were accommodated within an internal trough in the mandible [see (*2*)]. From dental evidence, *Teinolophos* is unequivocally placed in the monotreme lineage (*2*, *10*). But like *Morganucodon* and very much unlike living monotremes, *Teinolophos* exhibits a well-developed internal mandibular trough, suggesting that the angular (tympanic), the articular (malleus), and other "reptilian" jaw bones remained attached to the mandible through ligaments long after *Teinolophos* and living monotremes split from the common ancestor of marsupials and placentals. Another recent study shows that the middle ear bones were no longer accommodated by the internal mandibular trough but were still linked via the ossified Meckel's cartilage to the mandible in some triconodont mammals. These mammals evolved after the divergence of the monotreme and therian (marsupial and

sensing the path of a current in a conductor or in microfabricated heterostructures.

Another limit to miniaturization is set by the electrostatic attraction between atoms that are adsorbed and polarized at the chip surface. These forces become effective at micrometer-scale distances. A final limit is set by weak, attractive dispersion forces between the chip surface and the atom cloud. They rapidly set in at distances of several hundred nanometers and are difficult to compensate because of their strong spatial dependence.

Given these limits, there are several directions for future developments. Microfabrication can be improved to reduce imperfections and obtain more homogeneous current conductors—for example, by switching from electroplating to direct vapor deposition. However, basic atom chips can already be created with existing technology. Spatial fluctuations of the magnetic field due to imperfections of the conductors average out over distances that are large relative to the size of the imperfections. A waveguide far from the chip surface exhibits a smooth potential. This smooth waveguide can be combined with steep and thin potential barriers, generated on an extra chip next to the atomic cloud. Such a geometry, in which waveguide and potential barriers are generated by separate chips, is far less sensitive to wire imperfections than a single-layer chip. Matter-wave interferences have recently been observed with such a setup (*10*).

Atomic matter waves on chips can also be manipulated with electric fields (*11*) or even laser beams (*12*). Both induce an electric dipole moment through which the interaction with atoms is achieved. Combinations of magnetic, electrostatic, and optical potentials on atom chips remain to be explored. For example, temporally controlled lasers can be used to produce potential barriers in magnetic waveguides.

Instead of manipulating a condensate as a whole, it can also be used as a zero-temperature reservoir for single atoms. Such an atom laser could provide a coherent source for performing optics with single atoms. To measure and detect single atoms after they have been manipulated, a detector with single-atom sensitivity would be desirable. Possible approaches include optical ionization with light from the tip of an optical fiber followed by ion detection, and detection of single atoms in miniaturized optical resonators. According to theoretical studies, present-day technology should permit the construction of microcavities that can detect a single atom in the trap (*13*).

Which of the various ideas will prove most successful remains to be shown. As in every young field of research, atom chips will be good for future surprises.

References

1. J. D. Weinstein, K. G. Libbrecht, *Phys. Rev. A* **52**, 4004 (1995).
2. H. Ott, J. Fortágh, G. Schlotterbeck, A. Grossmann, C. Zimmermann, *Phys. Rev. Lett.* **87**, 230401 (2001).
3. W. Hänsel, P. Hommelhoff, T. W. Hänsch, J. Reichel, *Nature* **413**, 498 (2001).
4. W. Hänsel *et al.*, *Phys. Rev. A* **64**, 063607 (2001).
5. A. Smerzi *et al.*, *Phys. Rev. Lett.* **79**, 4950 (1997).
6. U. Dorner *et al.*, *Phys. Rev. Lett.* **91**, 073601 (2003).
7. J. Fortágh *et al.*, *Opt. Commun.* **243**, 45 (2004).
8. C. Henkel, S. Pötting, M. Wilkens, *Appl. Phys. B* **69**, 379 (1999).
9. J. Estève *et al.*, *Phys. Rev. A* **70**, 043629 (2004).
10. J. Fortágh, Institute for Theoretical Atomic, Molecular and Optical Physics (ITAMP) Workshop on Quantum Degenerate Gases in Low-Dimensionality, Cambridge, MA, 4 to 6 October 2004 (http://itamp.harvard.edu/low-d/low-d.html).
11. P. Krüger *et al.*, *Phys. Rev. Lett.* **91**, 233201 (2003).
12. Y.-J. Wang *et al.*, http://arxiv.org/abs/cond-mat/0407689.
13. P. Horak *et al.*, *Phys. Rev. A* **67**, 043806 (2003).

10.1126/science.1107348

PALEONTOLOGY

Homoplasy in the Mammalian Ear

Thomas Martin and Zhe-Xi Luo

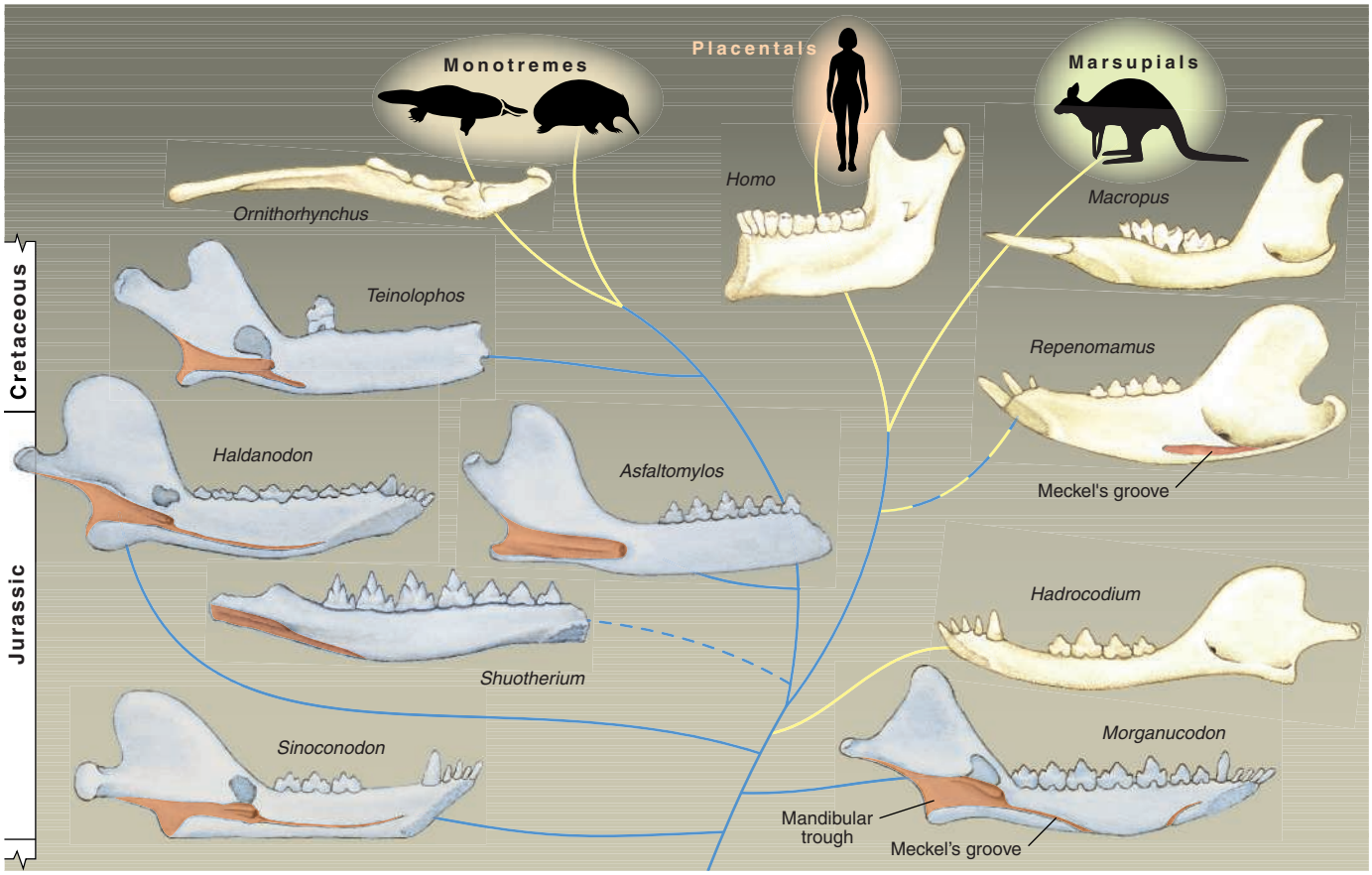
The similarity among structures that arose through independent evolution instead of descent from a common ancestor is termed homoplasy and is a major feature of evolutionary morphology. A fascinating but very difficult question facing evolutionary biologists is whether a complex structure would be less likely than a simple structure to undergo independent homoplastic evolution (*1*). On page 910 of this issue, Rich *et al*. (*2*) partially answer this question with their analysis of the dentary bone from the lower jaw of an Early Cretaceous fossil monotreme called *Teinolophos*, an extinct relative of Australia's modern platypus and echidna. The new fossil find offers fresh anatomical evidence to support the hypothesis that a key evolutionary innovation among modern mammals—the separation of the middle ear bones from the mandible—must have evolved independently among the monotreme mammals and the therians (marsupials and placentals).

T. Martin is at the Forschungsinstitut Senckenberg, D-60325 Frankfurt am Main, Germany. E-mail: tmartin@senckenberg.de Z.-X. Luo is at the Carnegie Museum of Natural History, Pittsburgh, PA 15213, USA. E-mail: luoz@carnegiemnh.org

The tiny bones of the middle ear that are used for hearing render modern mammals—including placentals, pouched marsupials, and egg-laying monotremes—unique among vertebrates (*2*). The middle ear bones are the malleus, incus, and stapes, and in addition there is the tympanic bone, which supports the tympanic membrane, enabling it to receive sound. The tympanic, malleus, and incus are homologous to bones in the mandible and jaw hinge (the angular, articular, and quadrate, respectively), which are required for feeding in nonmammalian vertebrates (*3*–*5*). There is also extensive evidence from fossils of extinct cynodont and mammaliaform relatives of modern mammals suggesting that the angular, articular, and quadrate bones in these creatures were used for hearing while still attached to the mandible and jaw hinge (*6*, *7*). Evolution of the mammalian jaw joint and middle ear represents a classic example of the phylogenetic transformation of a complex functional structure that can be read directly from fossil evidence.

However, alternative interpretations have waxed and waned about how these middle ear bones got separated from the mandible during early mammalian evolution. The structure of the middle ear is so complex and unique that some researchers consider the separation of the middle ear bones from the mandible to be the strongest synapomorphic (shared derived) characteristic of living mammals (*7*, *8*). They also propose that the monotreme, marsupial, and placental lineages split after their common ancestor had acquired this key feature. This view has been contested by others who favor multiple and independent acquisitions of the mammalian middle ear bones after the divergence of monotremes, marsupials, and placentals (*6*, *9*).

In premammalian cynodonts and such primitive mammaliaforms as *Morganucodon* (see the figure), the middle ear bones were accommodated within an internal trough in the mandible [see (*2*)]. From dental evidence, *Teinolophos* is unequivocally placed in the monotreme lineage (*2*, *10*). But like *Morganucodon* and very much unlike living monotremes, *Teinolophos* exhibits a well-developed internal mandibular trough, suggesting that the angular (tympanic), the articular (malleus), and other "reptilian" jaw bones remained attached to the mandible through ligaments long after *Teinolophos* and living monotremes split from the common ancestor of marsupials and placentals. Another recent study shows that the middle ear bones were no longer accommodated by the internal mandibular trough but were still linked via the ossified Meckel's cartilage to the mandible in some triconodont mammals. These mammals evolved after the divergence of the monotreme and therian (marsupial and



From feeding to hearing. Convergent separation of the middle ear bones from the mandible in the monotreme and therian (marsupial and placental) mammalian lineages. Convergent separation is inferred from the independent loss of the mandibular structures for the attachment of the middle ear bones. Primitive mammalian lineages (gray; blue lines) have a plesiomorphic mandibular trough for full accommodation and attachment of the middle ear bones to the mandible. The internal mandibular ("postdentary") trough is shown in brown. Derived lineages (cream; yellow lines) have lost the mandibular trough and show final separation of the middle ear bones from the mandible. Middle ear bones that are no longer accommodated by the mandibular trough but are still linked by the Meckel's cartilage to the middle of the mandible are indicated by the yellow and blue dashed line (*11*). [Tree topology is based on (*12*–*14*, *16*); figure is not to scale.]

placental) lineages (*11*). The oldest fossil lacking an internal mandibular trough is the Early Jurassic mammaliaform *Hadrocodium* (see the figure) (*12*), which suggests that detachment of the middle ear bones from the mandible occurred among extinct mammaliaforms during the Early Jurassic and before the diversification of modern mammalian clades in the Middle Jurassic (*13*, *14*). No matter how complex the modern mammalian middle ear seems to be, when its distribution is mapped onto the latest mammalian family tree, the middle ear bones appear to have evolved at least three times (see the figure).

This revelation about the convergent evolution of middle ear bones is certainly exciting enough for most evolutionary morphologists. But even more dramatic is the fact that the new *Teinolophos* fossil bears an uncanny resemblance to the australosphenidan mammals of Australia (*15*) and South America (*16*, *17*). These creatures also possess a very similar internal mandibular trough, indicating that the persisting attachment of the middle ear bones to the mandible is a common feature of both the monotremes and the australosphenidan mammals with tribosphenic molars from Gondwanaland (southern continents) from the Middle Jurassic to the Early Cretaceous, which contrasts sharply with their highly advanced molars. Given current morphological evidence, parsimony predicts the odds in favor of the homoplastic evolution of both the middle ear bones and derived molar features of the australosphenidan mammals (thought by some to include monotremes) (*13*, *16*).

Recent discoveries of Mesozoic mammals from Gondwanaland, including the toothed monotreme *Teinolophos*, have already changed our perspective of early mammalian evolution (*2*, *10*, *13*–*17*). Teeth and jaws of *Teinolophos* and its relatives offer a mere hint of how little we know about the diversity of the endemic Mesozoic mammals of Gondwanaland. The new fossil find described by Rich and colleagues sheds light on the great complexity of mammalian genealogical history. It also requires us to accommodate the homoplastic evolution of many complex mammalian cranial and dental structures that would have been unthinkable when we had only the fossil record of Laurasia to consider.

References

1. G. Wesley-Hunt, *J. Vertebr. Paleontol.* **24** (suppl. 3), 128A (2004).
2. T. H. Rich, J. A. Hopson, A. M. Musser, T. F. Flannery, P. Vickers-Rich, *Science* **307**, 910 (2005).
3. C. Reichert, *Müllers Arch. Anat. Physiol. Wiss. Med.* **1837**, 120 (1837).
4. E. Gaupp, *Archiv Anat. Entwickl.* **1912**, 1 (1913).
5. M. Sánchez-Villagra, S. Gemballa, S. Nummela, K. K. Smith, W. Maier, *J. Morphol.* **251**, 219 (2002).
6. E. F. Allin, J. A. Hopson, in *Evolutionary Biology of Hearing*, D. G. Webster, R. R. Fay, A. N. Popper, Eds. (Springer-Verlag, New York, 1992), pp. 587–614.
7. T. Rowe, *Science* **273**, 651 (1996).
8. T. S. Kemp, *Zool. J. Linn. Soc.* **77**, 353 (1983).
9. G. Fleischer, *Adv. Anat. Embryol. Cell Biol.* **55**, 3 (1978).
10. T. Rich *et al.*, *Acta Palaeo. Pol.* **46**, 113 (2001).
11. Y.-Q. Wang, Y.-M. Hu, J. Meng, C.-K. Li, *Science* **294**, 357 (2001).
12. Z.-X. Luo, A. W. Crompton, A.-L. Sun, *Science* **292**, 1535 (2001).
13. Z.-X. Luo, R. Cifelli, Z. Kielan-Jaworowska, *Nature* **409**, 53 (2001).
14. Z. Kielan-Jaworowska, R. Cifelli, Z.-X. Luo, *Mammals from the Age of Dinosaurs, Origins, Evolution and Structure* (Columbia Univ. Press, New York, 2004).
15. T. H. Rich *et al.*, *Science* **278**, 1438 (1997).
16. O. W. M. Rauhut, T. Martin, E. Ortiz-Jaureguizar, P. Puerta, *Nature* **416**, 165 (2002).
17. A. Forasiepi, G. Rougier, A. Martinelli, *J. Vertebr. Paleontol.* **24** (suppl. 3), 59A (2004).

10.1126/science.1107202

CREDIT: EMESE KAZÁR

YOUNG SCIENTIST AWARD

Construction of a Minimal, Protein-Free Spliceosome

Saba Valadkhan

Splicing of pre-messenger RNAs to mature transcripts is a crucial step in eukaryotic gene expression. Almost all primary transcripts in higher eukaryotes undergo multiple splicing events, and alternative splicing plays a major role in establishing proteomic diversity in higher eukaryotes. There are numerous examples where splicing and its regulation play key roles in cell growth control, differentiation, and disease. Consistent with its critical role, the spliceosome (the massive ribonucleoprotein particle that catalyzes splicing) has been shown to be the largest and most complicated molecular machine known (*1*).

The complexity of the spliceosome stems not only from its large number of components—almost 300 proteins and five small nuclear RNAs (snRNAs)—but also from the fact that the spliceosome is dynamic, assembling for each splicing event in an elaborate and stepwise fashion involving multiple rearrangements and conformational changes. This complexity poses severe limitations on experimental approaches that can be used to study its function; thus, despite intense research, fundamental aspects of the spliceosome function, such as identity of the catalytic domains and organization of the active site, have remained elusive.

At the start of my graduate research, the existing data pointed to two of the spliceosomal snRNAs, U2 and U6, as the likely catalytic players in the spliceosome. The presence of invariant, functionally crucial domains in these snRNAs, and their similarity to self-splicing group II introns (natural ribozymes that catalyze a reaction identical to the splicing reaction), made them likely candidates for the spliceosomal catalytic domain (*2*). However, despite two decades of experimental effort, neither the catalytic competence of the snRNAs nor direct involvement of other spliceosomal components in catalysis had been established (*3*).

In my graduate work, I addressed this question by trying to build the spliceosomal active site from scratch, by putting together the catalytically essential parts of U2 and U6 and adding short RNA oligonucleotides as splicing substrates. Developing a minimal active site would allow me to test directly the catalytic potential of these two snRNAs in isolation, thereby addressing the question of catalytic domain identity and the possibility of RNA catalysis in the spliceosome. It would also provide a much-needed minimal splicing system for studying the mechanistic aspects of splicing not amenable to study in the context of the whole spliceosome.

The author is at Case Western University, 2109 Adelbert Road, Wood RT 100-8, Cleveland, OH 44106, USA. E-mail: saba.valadkhan@case.edu

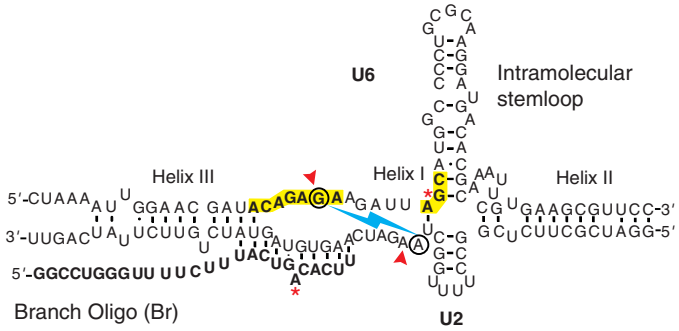
YOUNG SCIENTIST AWARD

GE Healthcare and *Science*/AAAS are pleased to present the prize-winning essay by Saba Valadkhan, a regional winner from North America, who is the Grand Prize winner of the Young Scientist Award.

To this end, I first showed that U2 and U6 snRNAs have the intrinsic ability to interact with each other in a manner similar to that observed in the active spliceosome. In the absence of any protein factor, U2 and U6 formed a base-paired scaffold that brought the catalytically crucial regions of the two molecules close to each other, and the presence of a functionally important tertiary fold, previously observed in the authentic spliceosome, helped to further position the catalytic domains. Thus, it seemed that the ability to create at least a partial spliceosomal active site was inherent in these two snRNAs (*4*).

Next, we assayed the ability of the U2–U6 complex to recognize and correctly bind the splicing substrates. Two short RNA oligonucleotides, one carrying the consensus sequence for the 5′ splice site (5′SS), the other containing the branch site consensus (Br), were added to the system. Although U2–U6 could efficiently bind Br through base pairing, the 5′SS oligonucleotide, which could form fewer base pairs, was not recognized. Nevertheless, we assayed the system for catalytic activity in the presence of these two oligonucleotides. Interestingly, we obtained an RNA species resulting from a specific catalytic activity (*5*, *6*). Characterization of the RNA, which we named RNA X, showed that its formation depended on intact U2, U6, and Br. However, the 5′SS oligonucleotide was not needed for this reaction. Further characterization of RNA X proved that it resulted from a splicing-related reaction that showed identical sequence and ionic requirements to the authentic splicing reaction. Accordingly, our protein-free reaction and splicing were likely catalyzed by the same, or a very similar, active site. In sum, the protein-free snRNAs were shown to have splicing-related catalytic activity.

The challenge then became finding a way to make the system correctly bind and position the 5′SS, thus setting the conditions for the authentic splicing reaction to occur. It has been shown that, in nematode spliceosomes, certain U6 mutations lead to erroneous use of the 5′ end of U6 in place of the 5′ splice site substrate in splicing reactions (*7*). Since this observation proved it permissible to put the 5′ splice site and U6 on the same molecule, we covalently attached 5′SS to the 5′ end of U6,



The spliceosome active site. The in vitro–assembled, protein-free U2/U6 complex. Yellow highlighted boxes mark the invariant regions in U6, and previously established base-paired regions are indicated. The circled residues connected by thunderbolts can be cross-linked by UV light. Arrows point to residues involved in a genetically proven interaction in yeast (*8*). The base-pairing interactions between Br and U2 snRNA are indicated. The asterisks denote the residues involved in the covalent link between Br and U6 in a splicing-related reaction catalyzed by U2 and U6.

Published by AAAS

with a linker sequence joining the two. We engineered a hyperstable hairpin in the linker region to orient 5′SS toward the active site and adjusted the length and sequence of the rest of the linker to help position 5′SS in register with critical active-site nucleotides. Using this chimeric construct, we screened for catalytic products that not only required U2, U6, 5′SS, and Br for their formation but also contained a chemical linkage identical to the product of the first splicing step. One candidate (RNA Y) seemed to satisfy all these requirements. Not only are all the correct sequence elements required for its formation, but our analysis also showed that the chemistry of the reaction was identical to that of the first step of splicing. These studies collectively proved that the spliceosome is an RNA enzyme and a relic from the RNA world.

In addition to providing direct evidence for RNA catalysis in the spliceosome and thus settling the longstanding and central question of identity of the catalytic domain, the minimal system provides a powerful tool for studying the spliceosome. Moreover, the minimal system can be compared with group II introns and the authentic spliceosome in attempts to understand the evolutionary origin of the spliceosome and the transition of the RNA world to the modern, protein-dominated one. Taken together, in addition to putting the spliceosome on the growing list of RNA enzymes playing central roles in cellular function, our results provide an example of the use of designed simple biological systems to address problems not amenable to other approaches.

References and Notes

1. T. W. Nilsen, *Bioessays* **25**, 1147 (2003).
2. S. Valadkhan, J. L. Manley, *Nature Struct. Biol.* **9**, 498 (2002).
3. C. A. Collins, C. Guthrie, *Nature Struct. Biol.* **7**, 850 (2000).
4. S. Valadkhan, J. L. Manley, *RNA* **6**, 206 (2000).
5. S. Valadkhan, J. L. Manley, *Nature* **413**, 701 (2001).
6. S. Valadkhan, J. L. Manley, *RNA* **9**, 892 (2003).
7. Y. T. Yu, P. A. Maroney, T. W. Nilsen, *Cell* **75**, 1049 (1993).
8. H. D. Madhani, C. Guthrie, *Genes Dev.* **8**, 1071 (1994).

10.1126/science.1110022

2004 Grand Prize Winner

Saba Valadkhan was born and raised in Tehran, Iran. She attended medical school at the Iran University of Medical Sciences from 1989 to 1996 and in 1993 placed fourth in the country in the nationwide Basic Sciences Medical Board Exam. She moved to the United States in 1996 to attend graduate school at Columbia University, New York. There she studied the role of small nuclear RNAs in the human spliceosome under the supervision of Prof. James Manley. While at Columbia University she received awards for both teaching and research. Her thesis was recognized with a Harold Weintraub award from the Fred Hutchinson Cancer Research Center in Seattle. In 2004 she joined Case Western Reserve University in Cleveland, Ohio, as an assistant professor and was named a Searle Scholar the same year.

Regional Winners

North America: Benjamin Tu for his essay, "Deciphering Disulfide Bonds." Dr. Tu was born in Stanford, California, but grew up in Pennsylvania. After graduating from Harvard in 1998, he moved to the University of California, San Francisco. He was awarded a Howard Hughes Medical Institute Predoctoral Fellowship and joined the laboratory of Dr. Jonathan S. Weissman, where he worked on a long-term problem in protein folding—how disulfide bonds are formed in proteins that traverse the secretory pathway. Dr. Tu obtained his Ph.D. in 2003 and moved to the University of Texas Southwestern Medical Center in Dallas, where he is a postdoctoral fellow in the laboratory of Dr. Steven L. McKnight with a fellowship from the Helen Hay Whitney Foundation. He is currently studying the metabolic cycles of yeast and hopes to apply what he learns to the study of circadian rhythms. In his spare time, he enjoys ultimate frisbee, tennis, ping-pong, puzzles, and Starcraft.

Europe: Christian Haering for his essay, "A Ring for Holding Sister Chromatids Together?" Dr. Haering grew up in Bavaria, Germany. He graduated from the University of Regensburg, Germany, with a diploma in biochemistry in 1999 and in 2000 joined Prof. Kim Nasmyth's group at the Research Institute of Molecular Pathology (IMP) in Vienna, Austria. Dr. Haering showed that the cohesin complex, required for proper chromosome segregation during cell division, forms a large ring structure with the potential to hold sister chromatids together by trapping them inside rings. His work was recognized with the Austrian Cell Cycle Publication Award in 2002. He received his Ph.D. in 2003 and has recently moved to the Department of Biochemistry at the University of Oxford, UK, where he plans to continue working on the mechanism of the cohesin complex.

Japan: Kunihiko Nishino for his essay, "Analysis of Drug Exporter Gene Libraries Based on Genome Information and Study of Their Regulatory Networks." Dr. Nishino was born in Kyoto, Japan. After graduating from Kyoto Pharmaceutical University he joined the Graduate School of Pharmaceutical Sciences at Osaka University in 1998. Working in the laboratory of Dr. Akihito Yamaguchi, he developed a postgenomic approach to understanding the regulation and function of xenobiotic exporters. He was awarded a Research Fellowship for Young Scientists from the Japan Society for the Promotion of Science in 2001 and the Kuroya Award from the Japanese Society for Bacteriology in 2002. After receiving his Ph.D. in 2003, he collaborated with Dr. Takeshi Honda at the Research Institute for Microbial Diseases, Osaka University, to extend his knowledge of bacterial pathogenesis. He is now a postdoctoral fellow in the laboratory of Dr. Eduardo A. Groisman at the Howard Hughes Medical Institute, Washington University School of Medicine, St. Louis, Missouri.

All Other Countries: Suvendra Bhattacharyya for his essay, "Mitochondrial tRNA Import: Glimpses of a Complex Molecular Machine." Dr. Bhattacharyya was born in Calcutta, India. He attended the University of Calcutta, where he obtained his B.S. in chemistry in 1996 and M.Sc. in biochemistry in 1998. He then joined the laboratory of Dr. Samit Adhya at the Indian Institute of Chemical Biology (IICB), where he purified the first mitochondrial RNA import complex (RIC) and established a model of tRNA import in *Leishmania* mitochondria. After completing his Ph.D. in 2003, he went to Basel, Switzerland, to join Prof. Witold Filipowicz's group at the Friedrich Miescher Institute (FMI), of the Novartis Research Foundation, with postdoctoral fellowships from the European Molecular Biology Organisation and the Human Frontier Science Program Organization. At FMI he is engaged in research on microRNA metabolism in mammalian cells. In 2004 he received the Young Scientist Award of the Indian National Science Academy (INSA), India.

For the full text of essays by the regional winners and for information about applying for next year's awards, see *Science* Online at www.sciencemag.org/feature/data/prizes/ge/index.shtml.

with a linker sequence joining the two. We engineered a hyperstable hairpin in the linker region to orient 5′SS toward the active site and adjusted the length and sequence of the rest of the linker to help position 5′SS in register with critical active-site nucleotides. Using this chimeric construct, we screened for catalytic products that not only required U2, U6, 5′SS, and Br for their formation but also contained a chemical linkage identical to the product of the first splicing step. One candidate (RNA Y) seemed to satisfy all these requirements. Not only are all the correct sequence elements required for its formation, but our analysis also showed that the chemistry of the reaction was identical to that of the first step of splicing. These studies collectively proved that the spliceosome is an RNA enzyme and a relic from the RNA world.

In addition to providing direct evidence for RNA catalysis in the spliceosome and thus settling the longstanding and central question of identity of the catalytic domain, the minimal system provides a powerful tool for studying the spliceosome. Moreover, the minimal system can be compared with group II introns and the authentic spliceosome in attempts to understand the evolutionary origin of the spliceosome and the transition of the RNA world to the modern, protein-dominated one. Taken together, in addition to putting the spliceosome on the growing list of RNA enzymes playing central roles in cellular function, our results provide an example of the use of designed simple biological systems to address problems not amenable to other approaches.

References and Notes

1. T. W. Nilsen, *Bioessays* **25**, 1147 (2003).
2. S. Valadkhan, J. L. Manley, *Nature Struct. Biol.* **9**, 498 (2002).
3. C. A. Collins, C. Guthrie, *Nature Struct. Biol.* **7**, 850 (2000).
4. S. Valadkhan, J. L. Manley, *RNA* **6**, 206 (2000).
5. S. Valadkhan, J. L. Manley, *Nature* **413**, 701 (2001).
6. S. Valadkhan, J. L. Manley, *RNA* **9**, 892 (2003).
7. Y. T. Yu, P. A. Maroney, T. W. Nilsen, *Cell* **75**, 1049 (1993).
8. H. D. Madhani, C. Guthrie, *Genes Dev.* **8**, 1071 (1994).

10.1126/science.1110022

2004 Grand Prize Winner

Saba Valadkhan was born and raised in Tehran, Iran. She attended medical school at the Iran University of Medical Sciences from 1989 to 1996 and in 1993 placed fourth in the country in the nationwide Basic Sciences Medical Board Exam. She moved to the United States in 1996 to attend graduate school at Columbia University, New York. There she studied the role of small nuclear RNAs in the human spliceosome under the supervision of Prof. James Manley. While at Columbia University she received awards for both teaching and research. Her thesis was recognized with a Harold Weintraub award from the Fred Hutchinson Cancer Research Center in Seattle. In 2004 she joined Case Western Reserve University in Cleveland, Ohio, as an assistant professor and was named a Searle Scholar the same year.

Regional Winners

North America: Benjamin Tu for his essay, "Deciphering Disulfide Bonds." Dr. Tu was born in Stanford, California, but grew up in Pennsylvania. After graduating from Harvard in 1998, he moved to the University of California, San Francisco. He was awarded a Howard Hughes Medical Institute Predoctoral Fellowship and joined the laboratory of Dr. Jonathan S. Weissman, where he worked on a long-term problem in protein folding—how disulfide bonds are formed in proteins that traverse the secretory pathway. Dr. Tu obtained his Ph.D. in 2003 and moved to the University of Texas Southwestern Medical Center in Dallas, where he is a postdoctoral fellow in the laboratory of Dr. Steven L. McKnight with a fellowship from the Helen Hay Whitney Foundation. He is currently studying the metabolic cycles of yeast and hopes to apply what he learns to the study of circadian rhythms. In his spare time, he enjoys ultimate frisbee, tennis, ping-pong, puzzles, and Starcraft.

Europe: Christian Haering for his essay, "A Ring for Holding Sister Chromatids Together?" Dr. Haering grew up in Bavaria, Germany. He graduated from the University of Regensburg, Germany, with a diploma in biochemistry in 1999 and in 2000 joined Prof. Kim Nasmyth's group at the Research Institute of Molecular Pathology (IMP) in Vienna, Austria. Dr. Haering showed that the cohesin complex, required for proper chromosome segregation during cell division, forms a large ring structure with the potential to hold sister chromatids together by trapping them inside rings. His work was recognized with the Austrian Cell Cycle Publication Award in 2002. He received his Ph.D. in 2003 and has recently moved to the Department of Biochemistry at the University of Oxford, UK, where he plans to continue working on the mechanism of the cohesin complex.

Japan: Kunihiko Nishino for his essay, "Analysis of Drug Exporter Gene Libraries Based on Genome Information and Study of Their Regulatory Networks." Dr. Nishino was born in Kyoto, Japan. After graduating from Kyoto Pharmaceutical University he joined the Graduate School of Pharmaceutical Sciences at Osaka University in 1998. Working in the laboratory of Dr. Akihito Yamaguchi, he developed a postgenomic approach to understanding the regulation and function of xenobiotic exporters. He was awarded a Research Fellowship for Young Scientists from the Japan Society for the Promotion of Science in 2001 and the Kuroya Award from the Japanese Society for Bacteriology in 2002. After receiving his Ph.D. in 2003, he collaborated with Dr. Takeshi Honda at the Research Institute for Microbial Diseases, Osaka University, to extend his knowledge of bacterial pathogenesis. He is now a postdoctoral fellow in the laboratory of Dr. Eduardo A. Groisman at the Howard Hughes Medical Institute, Washington University School of Medicine, St. Louis, Missouri.

All Other Countries: Suvendra Bhattacharyya for his essay, "Mitochondrial tRNA Import: Glimpses of a Complex Molecular Machine." Dr. Bhattacharyya was born in Calcutta, India. He attended the University of Calcutta, where he obtained his B.S. in chemistry in 1996 and M.Sc. in biochemistry in 1998. He then joined the laboratory of Dr. Samit Adhya at the Indian Institute of Chemical Biology (IICB), where he purified the first mitochondrial RNA import complex (RIC) and established a model of tRNA import in *Leishmania* mitochondria. After completing his Ph.D. in 2003, he went to Basel, Switzerland, to join Prof. Witold Filipowicz's group at the Friedrich Miescher Institute (FMI), of the Novartis Research Foundation, with postdoctoral fellowships from the European Molecular Biology Organisation and the Human Frontier Science Program Organization. At FMI he is engaged in research on microRNA metabolism in mammalian cells. In 2004 he received the Young Scientist Award of the Indian National Science Academy (INSA), India.

For the full text of essays by the regional winners and for information about applying for next year's awards, see *Science* Online at www.sciencemag.org/feature/data/prizes/ge/index.shtml.

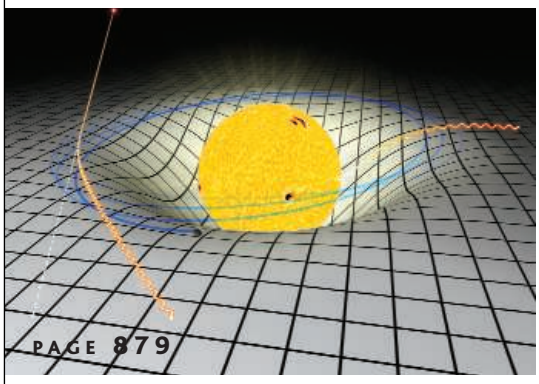
INTRODUCTION

A Passion for Physics

The World Year of Physics in 2005 celebrates the publication of Einstein's papers in *Annalen der Physik* that, 100 years ago, heralded the era of modern physics. Part of Einstein's legacy is that quantum mechanics and relativity forced us to accept notions of space and time, and of matter and energy, that do not square with our everyday experience. Given Einstein's passion for the deepest questions in physics, we highlight some of those challenges in this special issue.

Of Einstein's many contributions, the work that has most gripped the public imagination boils down to a single word: relativity. Today physicists are still wrestling with the implications of this mathematically straightforward but surreal-sounding theory. Some hope to find subtle flaws in a key component of the special theory of relativity—flaws that, as Cho explains in his News story (p. 866), could lead physics into realms even stranger than the ones that Einstein pioneered. In another News feature (p. 869), Seife and Lawler describe how NASA's new emphasis on exploration is affecting satellites designed to test whether the later general theory of relativity accurately describes spacetime.

Despite being one of the fathers of quantum mechanics, Einstein was dismayed by its probabilistic nature. Leggett (p. 871) explores the current status of the theory and discusses the controversy that still remains. Dunningham *et al*. (p. 872) look at two different categories: the fragile but useful and the robust but essentially useless. These two types of entanglement can perhaps explain the transition from the quantum to the classical worlds. Aharonov and Zubairy (p. 875) discuss some of the stranger aspects of time in quantum mechanics, especially the "delayed choice quantum eraser" of Scully and Drühl, in which the erasing of "which path" information in two-slit experiments can restore interference patterns, even if that information comes "after the fact."


PAGE 879

After the development of general relativity, Einstein spent many years attempting to unify the fundamental forces of the day: gravity and electromagnetism. Bennett (p. 879) discusses some of the current tests of Einstein's theories of relativity, including gravitational redshifts, gravitational waves, and the Lense-Thirring effect. He also discusses the role of dark energy and dark matter and the questions of the origin and ultimate fate of the universe. Guth and Kaiser (p. 884) overview inflationary cosmology, which helps bridge particle physics and gravitation, and they outline the experimental tests of inflation's mark on the present universe. They also discuss the theoretical efforts to use superstring theory to understand inflation and to determine the vacuum energy of the universe. In Books *et al*., Wilczek (p. 852) reviews Penrose's comprehensive perspective on our understanding of the physical behavior of the universe and the mathematical theory that underlies it.

Einstein's legacy includes the physicists who continue to take up his mantle. *Science*'s Next Wave (www.nextwave.org) profiles some European physicists as well as Canada's Perimeter Institute for Theoretical Physics.

–ROBERT COONTZ, IAN OSBORNE, AND PHIL SZUROMI

CONTENTS

See also related Next Wave material on p. 809 and Book Review on p. 852.

CREDIT: BENNETT *ET AL.*

Published by AAAS

NEWS

Special Relativity Reconsidered

Einstein's special theory of relativity reaches into every corner of modern physics. So why are so many trying so hard to prove it wrong?

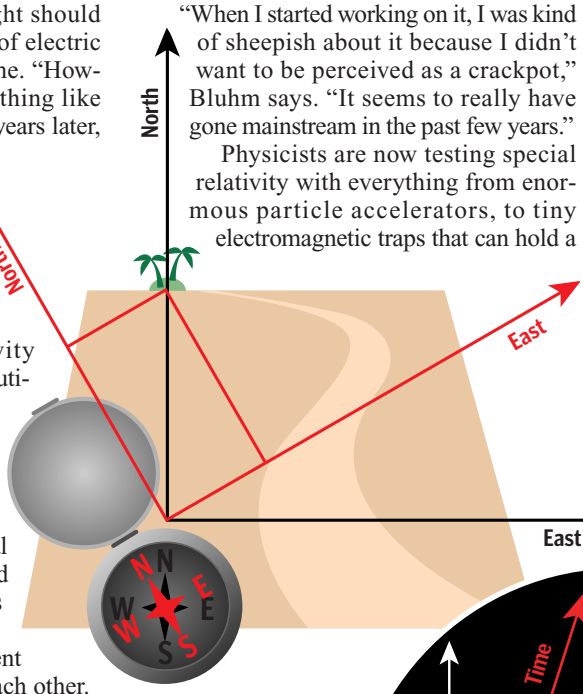
At an age when most boys would rather chase girls, Albert Einstein daydreamed of chasing light. When he was about 16 years old, Einstein later recalled, he realized that if he ran fast enough to catch up to it, light should appear to him as a wavy pattern of electric and magnetic fields frozen in time. "However," Einstein observed, "something like that does not seem to exist!" Ten years later, that insight blossomed into the special theory of relativity, which forbade catching light, overturned ancient conceptions of time and space, and laid the cornerstone for modern physics. Now, however, some physicists wonder whether special relativity might be subtly—and perhaps beautifully—wrong.

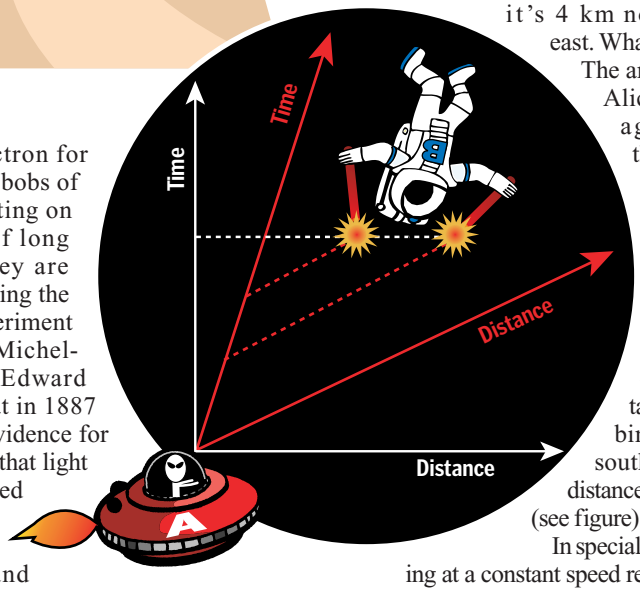
In 1905, physicists believed space was a grand stage on which the drama of the universe unfolded and time ticked away at the same rate for all actors. Special relativity denied all that. It replaced space and time as distinct entities with a single "spacetime" that, in mind-bending ways, looks different to observers moving relative to each other. But the theory's implications reach far beyond questions of when and where. Combined with quantum mechanics, it helps explain the stability of matter and even requires the existence of antimatter, says Steven Weinberg, a theoretical physicist at the University of Texas, Austin. "That's the only way nature can be if you're going to satisfy the requirements of both relativity and quantum mechanics," Weinberg says.

Yet a growing number of physicists are entertaining the possibility that special relativity is not quite correct. That may sound perverse, but researchers have good reason to hope Einstein's theory isn't the final word: Any deviation from special relativity could point physicists toward an elusive goal, a quantum theory of gravity. Candidate theories can be tested directly only with particle collisions a million, billion times more energetic than any produced with a particle accelerator. On the other hand, testing special relativity provides a far more practical, albeit indirect, way of probing quantum gravity, says V. Alan Kostelecký, a theorist at Indian University, Bloomington.

Only a decade ago, questioning special relativity would have struck many as heretical, says Robert Bluhm, a theoretical physicist at Colby College in Waterville, Maine. "When I started working on it, I was kind of sheepish about it because I didn't want to be perceived as a crackpot," Bluhm says. "It seems to really have gone mainstream in the past few years."

Physicists are now testing special relativity with everything from enormous particle accelerators, to tiny electromagnetic traps that can hold a single electron for months, to bobs of metal twisting on the ends of long fibers. They are even repeating the famed experiment by Albert Michelson and Edward Morley that in 1887 found no evidence for the "ether" that light was supposed to ripple through just as sound ripples through air. In spite of these efforts, special relativity remains inviolate—so far.





Unbearable coincidences

According to legend, Einstein invented special relativity to explain the Michelson-Morley experiment. In truth, he worried more about conceptual puzzles in the theory of electricity and magnetism, which had been hammered out in the 1860s by the Scottish physicist James Clerk Maxwell, says Michel Janssen, a historian of science at the University of Minnesota, Twin Cities.

Consider the simplest electrical generator—a loop of wire and a magnet moving toward each other at constant speed. Current will flow through the wire. According to Maxwell's theory, different mechanisms drive the current depending on whether the wire is stationary in the presupposed ether and the magnet is moving or the other way around. Yet the current is the same in either case. That supposed coincidence was too fantastic for Einstein. "The idea that we would be dealing here with two fundamentally different situations was unbearable to me," he later wrote. But Einstein found he could show that the two cases were different ways of looking at the same thing—if he abandoned the ether and familiar notions of space and time.

A simple analogy captures the essence of Einstein's insight. Imagine two explorers, Alice and Bob, lost in a vast desert. From the top of a dune they spot an oasis. Alice pulls out her compass and range finder and determines that the oasis is 5 kilometers due north. Bob takes his own measurements and finds that it's 4 km north and 3 km east. What's gone wrong? The answer is simple: Alice and Bob disagree because their compasses don't line up. Each has a different notion of north, so what Alice takes to be a purely north-south distance, Bob takes to be a combination of north-south and east-west distances, and vice versa (see figure).

In special relativity, traveling at a constant speed relative to another observer mixes time and space in much the same way. For example, imagine that instead of explorers, Alice and Bob are astronauts in deep space. Suppose, in a fit of foolishness, Bob holds up a firecracker in each of his outstretched hands. He sets the explosives off as Alice zooms past at half the speed of light. If Bob sees both firecrackers flash at the same time, Alice will see them flash at different times. So what Bob perceives as a purely spatial distance, Alice perceives as a spatial

CREDITS: J. MOGLIA/*SCIENCE*

Published by AAAS

distance and a time interval. In essence, there is only one spacetime, which they perceive as different combinations of space and time.

In precisely the same way, in Einstein's analysis there is only one underlying "electromagnetic" field that requires no ether, and different observers slice it into different combinations of electric and magnetic fields. The underlying unity explains why the current in the simple generator is the same regardless of whether the magnet or the wire is moving. In fact, according to special relativity it's meaningless to say which is "really" moving.

Outgrowing Einstein

Einstein doggedly followed his theory to bizarre but unavoidable conclusions. A clock whizzing by at near light speed ticks slower than the watch on your wrist. A meter stick flying past looks shorter than one in your hands. Light travels at the same speed for all observers—so it cannot be caught.

But special relativity packs even more punch when combined with quantum mechanics to form "relativistic quantum field theory." That amalgam predicts the existence of antimatter and demands a kind of mirrorlike correspondence between matter and antimatter, which is known as CPT symmetry. It also forges a connection between how much particles spin like little tops and whether two or more of them can occupy the exact same quantum state. That "spin-statistics connection" explains why atoms and nuclei do not implode.

Antimatter must exist because quantum mechanics blurs notions of before and after, at least for particles, says the University of Texas's Weinberg. Suppose Alice throws an electron and Bob catches it. Observers moving at different speeds will disagree on how long it takes the electron to make the trip, but, sans quantum mechanics, all will agree that Alice throws it before Bob catches it. Mix in the uncertainty principle, however, and some observers will see Bob receive the electron before Alice tosses it, Weinberg says. "And the way that relativistic field theory gets around that difficulty is by reinterpreting it as Bob emits an antielectron that Alice receives," he says. However, the conceptual fix-up works only if the electron and antielectron have exactly the same mass and other properties—collectively, CPT symmetry.

The spin-statistics connection is less intuitive. All particles behave like little tops and can have only certain amounts of spin. For example, the photons that make up light have exactly one quantum of spin, whereas the electrons, protons, and neutrons that make up atoms have half a quantum. The spin-statistics connection says that no two identical particles can occupy the same quantum state if they have spin 1/2 (or 3/2, 5/2, etc.) That means the electrons in an atom cannot collapse into a tiny knot. Instead they stack into shell-like orbitals, and this arrangement keeps the atom stable. And it's a consequence of special relativity, says O. W. Greenberg, a theorist at the University of Maryland, College Park. "Violating the spin-statistics connection in a relativistic theory is, so far as we know, just impossible," Greenberg says.

Ironically, Einstein disdained the marriage of special relativity and quantum mechanics. "I know from experience how difficult it was to discuss quantum field theory with him," wrote his scientific biographer, Abraham Pais, who died in 2000. "Relativistic quantum field theory was repugnant to him." But Greenberg says we should not disparage Einstein because he didn't fully appreciate the implications of his own discovery: "It sort of outgrew Einstein."

As the world turns

Now, however, some physicists are hoping to reach beyond special relativity. Researchers generally agree that the ultimate theory of gravity cannot be a quantum field theory. Such theories assume that particles are infinitesimal points and spacetime is smooth. But according to the uncertainty principle, at minuscule scales spacetime ought to erupt into a chaotic froth that overwhelms any theory of point particles. On the other hand, calculations suggest that alternative theories—such as string theory, which assumes that every particle is really a tiny vibrating string—might not completely jibe with special relativity.

Unfortunately, sketchy quantum gravity theories cannot tell experimenters precisely what signs to look for, says Indiana University's Kostelecký. So over the last 15 years, he and his colleagues have taken another tack. They start with the relativistic quantum field theory that explains all the particles seen so far, the so-called Standard Model. They add to it myriad "background fields" that lace empty space. These resemble an electromagnetic field in that each points in some direction. But whereas electromagnetic fields arise from charges and currents, the background fields are woven into the vacuum. Known as the Standard Model Extension (SME), this catch-all theory clashes with special relativity because each background field provides a universal benchmark with which to determine whether an object is moving or not, or at least which direction it's going.

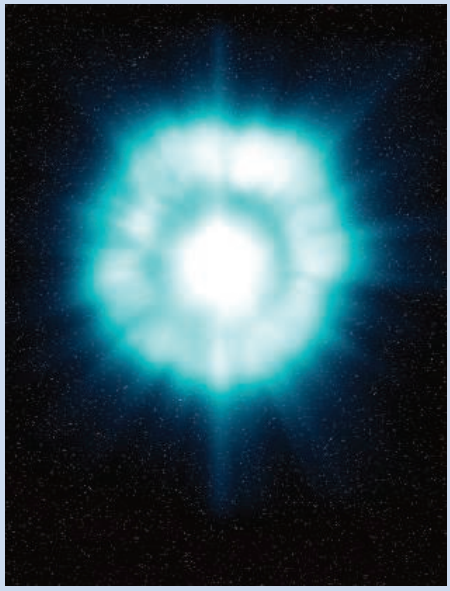
Experimenters are striving to glimpse the background fields, mainly by trying to detect Earth's motion through them. Because Earth spins, a lab will zoom through a background field at different angles at different times of

Doubly Special, Twice as Controversial

Quantum gravity may bend, not break, special relativity, some theorists say. Special relativity says that nothing can travel faster than light. Quantum gravity effects might also limit an individual particle's energy, says Giovanni Amelino-Camelia of the University of Rome "La Sapienza." That could lead to what he and others call "doubly special relativity." The embryonic theory has no background fields, and just as in ordinary special relativity, it's impossible to tell whether an object is moving relative to the vacuum. But the rules for adding up momentum and energy change, leading to potentially observable astronomical effects.

For example, doubly special relativity predicts that the speed of light could depend on its color and energy. Such an effect might be spotted by observing gargantuan stellar explosions known as gamma ray bursts, says Lee Smolin of the Perimeter Institute for Theoretical Physics in Waterloo, Canada. The gamma rays take billions of years to reach Earth, Smolin says, giving the faster ones time to pull measurably ahead of the slower ones.

However, some theorists doubt that doubly special relativity can be made into a coherent theory. Amelino-Camelia says he sees no obvious reason why it can't. Still, he adds, "there are plenty of consistency checks to be made, and I offer no guarantees until they're done." **–A.C.**

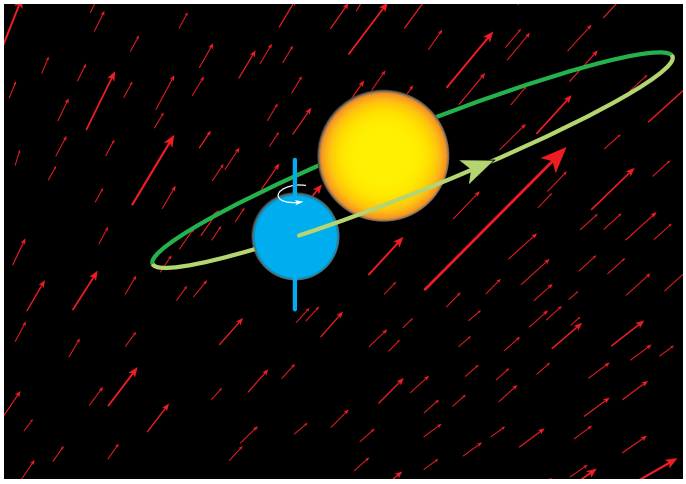


Photon photo finish. Gamma rays from humongous stellar explosions may reveal hypothesized variations in the speed of light.

ILLUSTRATION: ESA/ECF

Published by AAAS

day. So if an experiment bolted to the floor can feel the field, its output should oscillate in sync with Earth's rotation. Researchers expect the effects to be tiny. Spotting them will require, for example, measuring the frequencies of microwaves to one part in 1000 billion or better. But "even though these effects are very small," Kostelecký says, "the current experimental capabilities are in the range that you would expect to see something" if the fields are there.



Feel the breeze. Experimenters hope to detect oscillating effects as Earth whizzes and spins through putative background fields.

SME has limitations. It doesn't explain how the background fields arise. And each type of particle may interact with a different set of fields, leaving experimenters with dozens of measurements to make. Nevertheless, SME tells researchers which experiments should be most sensitive and enables them to compare seemingly disparate efforts, which is why it has sparked much of the interest in testing special relativity, says Blayne Heckel, an experimental physicist at the University of Washington, Seattle. "Once you have this community out there that appreciates what you're doing," Heckel says, "you don't feel so bad measuring zero."

K^0 mesons and clocks in space

Heckel is hardly the only one to come up short in trying to prove Einstein wrong: Experimenters have found no evidence that special relativity isn't bang on the money. For decades particle physicists have tested CPT symmetry, which now can be analyzed in the context of SME, by comparing particles and antiparticles. Researchers working with the KTeV experiment at Fermi National Accelerator Laboratory in Batavia, Illinois, have shown that fleeting particles called K^0 (pronounced kay-zero) mesons have the same mass as their antiparticles to one part in a billion billion. By "weighing" individual particles in devices called Penning traps, researchers at the European particle physics laboratory, CERN, near Geneva, Switzerland, have shown that protons and antiprotons have the same mass to a part in 10 billion.

Others are probing for background fields by comparing extremely precise clocks. A background field may affect one clock differently from the other, in which case one will speed up and slow down relative to the other over the course of the day. In fact, the "clocks" can be two different frequencies of radiation emitted by the same atom. Using a device called a maser, Ronald Walsworth of the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, and colleagues compared two frequencies emitted by hydrogen atoms, which allowed them to probe for fields that might affect the lone proton at the center of the hydrogen atom. No sign of the background fields emerged.

Some have resorted to a centuries-old technique. In 1798 English physicist Henry Cavendish measured the strength of gravity by dangling a barbell-shaped bob on the end of a fiber and watching it twist as one end came close to a heavy object. Now, Heckel and colleagues at the University of Washington have employed a souped-up version of Cavendish's "torsion balance" to search for background fields that interact with an electron's spin. The bob is a symmetrical assemblage of pieces of magnets arranged so that a majority of electrons spin in one direction. Crucially, the bob has no net magnetism, so it won't interact with the inevitable stray magnetic fields. So far Heckel and colleagues have seen no unusual twists of their apparatus.

Physicists have also repeated the Michelson- Morley experiment. Michelson and Morley reasoned that because Earth spins in the light-carrying ether, to an earthbound observer light should travel at different speeds depending on whether it is zipping north-south or east-west. SME's background fields could produce similar effects. Michelson and Morley used light beams and mirrors; today researchers employ "resonators" that ring with microwaves much as bells ring with sound. Two identical resonators are arranged perpendicularly and researchers compare their frequencies, achieving 10 million times better sensitivity than the original experiment, says John Lipa of Stanford University in California. Just as Michelson and Morley caught no whiff of the ether, modern experimenters have found no trace of the SME background fields.

Researchers had planned to fly atomic clocks and resonators on the international space station, where they would have been far more sensitive than earthbound experiments. But NASA scuttled those plans last year when President George W. Bush set his sights on sending humans to Mars (*Science*, 30 January 2004, p. 615). "Basically, NASA shut down all the activities in physics on the space station," Lipa says. "That's politics."

A subtler beauty

Why are some physicists so keen to take on Einstein? Answers vary widely. Experimenters should test basic theoretical assumptions as rigorously as possible as a matter of principle, says Gerald Gabrielse of Harvard University, who led the efforts to compare protons and antiprotons. "If we were to find a violation," Gabrielse says, "the consequences of that would ricochet through physics, affecting our understanding of the structure of the universe in every way."

On the other hand, some particle theorists may be drawn to the matter because "there's not that much else to do," quips Roman Jackiw, a theoretical physicist at the Massachusetts Institute of Technology in Cambridge. Particle theorists have little fresh and challenging data to gnaw on, Jackiw says, although that should change when an accelerator known as the Large Hadron Collider powers up at CERN in 2007.

To Kostelecký, the architect of SME, the allure is aesthetic. Special relativity states that spacetime possesses a kind of perfect symmetry, like an infinite plane so featureless that it's impossible to tell where you are and which direction you're facing. In special relativity, the symmetry extends to time, too, so that space and time mix together into a single seamless whole. That "Lorentz symmetry" is so elegant most physicists assume it's true. But "nature's beauty is more subtle than that perfect symmetry," Kostelecký says. "For me it may make nature more beautiful if it is *almost* Lorentz symmetric."

That sentiment might have intrigued Einstein, who often drew inspiration from his own sense of the beauty of nature and of physical theories. Perhaps he would have followed the thought to deep new insights, just as he surfed an imaginary light wave to one of the most profound ideas ever conceived. **–ADRIAN CHO**

ILLUSTRATION: ADAPTED FROM BLUHM, KOSTELECKÝ, LANE, RUSSELL, *PHYSICAL REVIEW LETTERS* **88** (2002)

distance and a time interval. In essence, there is only one spacetime, which they perceive as different combinations of space and time.

In precisely the same way, in Einstein's analysis there is only one underlying "electromagnetic" field that requires no ether, and different observers slice it into different combinations of electric and magnetic fields. The underlying unity explains why the current in the simple generator is the same regardless of whether the magnet or the wire is moving. In fact, according to special relativity it's meaningless to say which is "really" moving.

Outgrowing Einstein

Einstein doggedly followed his theory to bizarre but unavoidable conclusions. A clock whizzing by at near light speed ticks slower than the watch on your wrist. A meter stick flying past looks shorter than one in your hands. Light travels at the same speed for all observers—so it cannot be caught.

But special relativity packs even more punch when combined with quantum mechanics to form "relativistic quantum field theory." That amalgam predicts the existence of antimatter and demands a kind of mirrorlike correspondence between matter and antimatter, which is known as CPT symmetry. It also forges a connection between how much particles spin like little tops and whether two or more of them can occupy the exact same quantum state. That "spin-statistics connection" explains why atoms and nuclei do not implode.

Antimatter must exist because quantum mechanics blurs notions of before and after, at least for particles, says the University of Texas's Weinberg. Suppose Alice throws an electron and Bob catches it. Observers moving at different speeds will disagree on how long it takes the electron to make the trip, but, sans quantum mechanics, all will agree that Alice throws it before Bob catches it. Mix in the uncertainty principle, however, and some observers will see Bob receive the electron before Alice tosses it, Weinberg says. "And the way that relativistic field theory gets around that difficulty is by reinterpreting it as Bob emits an antielectron that Alice receives," he says. However, the conceptual fix-up works only if the electron and antielectron have exactly the same mass and other properties—collectively, CPT symmetry.

The spin-statistics connection is less intuitive. All particles behave like little tops and can have only certain amounts of spin. For example, the photons that make up light have exactly one quantum of spin, whereas the electrons, protons, and neutrons that make up atoms have half a quantum. The spin-statistics connection says that no two identical particles can occupy the same quantum state if they have spin 1/2 (or 3/2, 5/2, etc.) That means the electrons in an atom cannot collapse into a tiny knot. Instead they stack into shell-like orbitals, and this arrangement keeps the atom stable. And it's a consequence of special relativity, says O. W. Greenberg, a theorist at the University of Maryland, College Park. "Violating the spin-statistics connection in a relativistic theory is, so far as we know, just impossible," Greenberg says.

Ironically, Einstein disdained the marriage of special relativity and quantum mechanics. "I know from experience how difficult it was to discuss quantum field theory with him," wrote his scientific biographer, Abraham Pais, who died in 2000. "Relativistic quantum field theory was repugnant to him." But Greenberg says we should not disparage Einstein because he didn't fully appreciate the implications of his own discovery: "It sort of outgrew Einstein."

As the world turns

Now, however, some physicists are hoping to reach beyond special relativity. Researchers generally agree that the ultimate theory of gravity cannot be a quantum field theory. Such theories assume that particles are infinitesimal points and spacetime is smooth. But according to the uncertainty principle, at minuscule scales spacetime ought to erupt into a chaotic froth that overwhelms any theory of point particles. On the other hand, calculations suggest that alternative theories—such as string theory, which assumes that every particle is really a tiny vibrating string—might not completely jibe with special relativity.

Unfortunately, sketchy quantum gravity theories cannot tell experimenters precisely what signs to look for, says Indiana University's Kostelecký. So over the last 15 years, he and his colleagues have taken another tack. They start with the relativistic quantum field theory that explains all the particles seen so far, the so-called Standard Model. They add to it myriad "background fields" that lace empty space. These resemble an electromagnetic field in that each points in some direction. But whereas electromagnetic fields arise from charges and currents, the background fields are woven into the vacuum. Known as the Standard Model Extension (SME), this catch-all theory clashes with special relativity because each background field provides a universal benchmark with which to determine whether an object is moving or not, or at least which direction it's going.

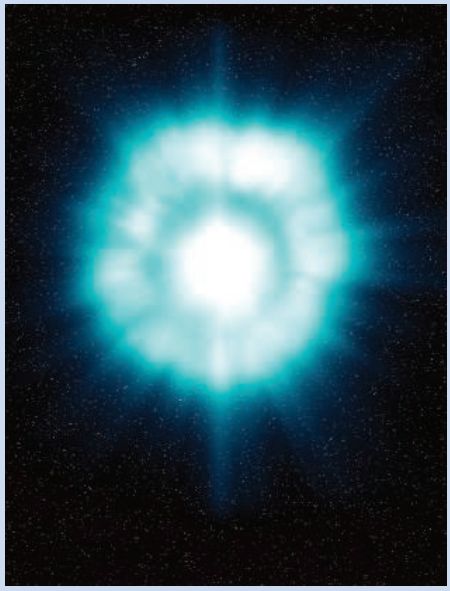
Experimenters are striving to glimpse the background fields, mainly by trying to detect Earth's motion through them. Because Earth spins, a lab will zoom through a background field at different angles at different times of

Doubly Special, Twice as Controversial

Quantum gravity may bend, not break, special relativity, some theorists say. Special relativity says that nothing can travel faster than light. Quantum gravity effects might also limit an individual particle's energy, says Giovanni Amelino-Camelia of the University of Rome "La Sapienza." That could lead to what he and others call "doubly special relativity." The embryonic theory has no background fields, and just as in ordinary special relativity, it's impossible to tell whether an object is moving relative to the vacuum. But the rules for adding up momentum and energy change, leading to potentially observable astronomical effects.

For example, doubly special relativity predicts that the speed of light could depend on its color and energy. Such an effect might be spotted by observing gargantuan stellar explosions known as gamma ray bursts, says Lee Smolin of the Perimeter Institute for Theoretical Physics in Waterloo, Canada. The gamma rays take billions of years to reach Earth, Smolin says, giving the faster ones time to pull measurably ahead of the slower ones.

However, some theorists doubt that doubly special relativity can be made into a coherent theory. Amelino-Camelia says he sees no obvious reason why it can't. Still, he adds, "there are plenty of consistency checks to be made, and I offer no guarantees until they're done." **–A.C.**



Photon photo finish. Gamma rays from humongous stellar explosions may reveal hypothesized variations in the speed of light.

ILLUSTRATION: ESA/ECF

Published by AAAS

NEWS

We're So Sorry, Uncle Albert

NASA's new focus on exploration closer to home may derail missions aimed at torture-testing Einstein's relativistic ideas

Einstein is in trouble. A century after his "miraculous year," astronomers and physicists across the globe have plotted an ambitious, multibillion-dollar challenge to Einstein's theory of relativity. Armadas of spacecraft launched over the next 2 decades will directly test some of the most dramatic assertions of relativity theory: that the entire fabric of space and time ripples with distortions, that there are regions in space where gravity is so strong that light cannot escape, and that the big bang and newly discovered "dark energy" leave a characteristic imprint upon the very distant and very ancient universe. Two great observatories, three smaller probes, and a pair of "vision missions"—which make up NASA's "Beyond Einstein" project—are the culmination of years of planning by astrophysicists.

The problem is not the tests. Most physicists believe that Einstein's theories will pass them handily and emerge strengthened by the new data. But when and whether the flotilla will be launched is now in question. When President George W. Bush announced last January that NASA would focus on lunar and Mars exploration by robots and humans, Beyond Einstein, which doesn't fit into that vision, faltered. The Administration cut budgets for parts of the effort and put others on the back burner.

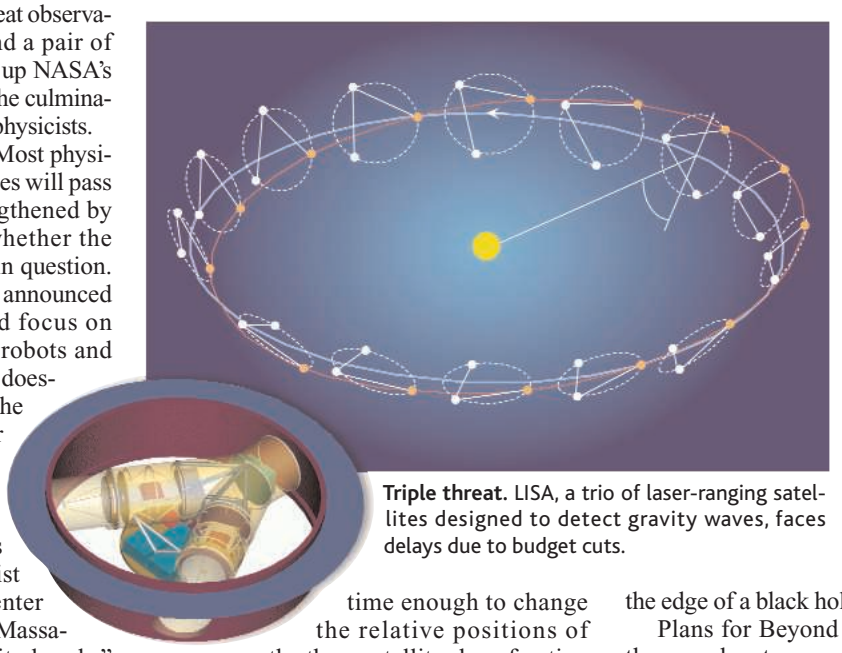
"I sincerely hope [Beyond Einstein] will survive," says Michael Garcia, an astrophysicist at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts. "I think it's taken a hit already." Unfortunately, Einstein's trouble may be that he's no longer an important part of NASA's universe.

Wrinkles and holes in time

NASA's Beyond Einstein effort tied together two existing projects called the Laser Interferometer Space Antenna (LISA) and Constellation-X, envisioned a new series of probes designed to answer fundamental questions raised by Einstein's work, and proposed innovative future missions to study black holes and peer back to the big bang. Launched with fanfare in February 2003, the program won congressional backing and the full $59 million NASA requested for 2004. It was an auspicious start for an effort estimated to cost $765 million during the first 5 years.

Beyond Einstein is aimed at addressing questions about Einstein's theory of relativity and about the nature of black holes and galaxies. LISA and Constellation-X, the two main observatories of the Beyond Einstein project, are both expensive flotillas of spacecraft, yet they attack those questions in very different ways.

LISA comprises three spacecraft that will surf the swells of spacetime, flying in formation—a 5-million-kilometer-wide triangle, linked by laser beams. As gravitational waves rattle by, they will stretch and squish space-time enough to change the relative positions of the three satellites by a fraction of a millimeter (*Science*, 16 August 2002, p. 1113). LISA will be able to see gravitational waves that earthbound observatories can't, in part because the vast distance between the satellites will make it much more sensitive.



Triple threat. LISA, a trio of laser-ranging satellites designed to detect gravity waves, faces delays due to budget cuts.

"The odds are in favor to see a substantial signal," says Peter Bender, a gravitational physicist at the Joint Institute for Laboratory Astrophysics in Boulder, Colorado. LISA should be able to pick up gravitational waves from various energetic events: the coalescence of massive black holes at the centers of galaxies, the inspirals of black hole binaries in the last moments before they collide, and perhaps even the ripples in space caused by a supernova explosion.

Detecting waves from these events will not only provide a ringing confirmation that Einstein's gravitational waves are real but will also tell astrophysicists about the nature of black holes and galaxy formation. "During the initial formation of galaxies, we think that they form hierarchically," says Bender—that small galaxies merge to form larger ones. "Whenever galaxies coalesce, you're likely to have their black holes coalescing also." So by listening to the ripples caused by crashing black holes, scientists would get a direct view of galaxy birth.

Constellation-X will also look at black holes, by sensing high-energy light. Matter close to a black hole is extraordinarily hot and emits x-rays; as a hunk of matter falls in, the immense gravitational field of the black hole stretches those x-rays and makes them redder and redder. "You can watch [a black hole's x-ray emissions] change through time," says Garcia. "You can watch it move from the blue to the red end of the spectrum." To spot those changes, scientists plan to yoke together four x-ray telescopes—their size limited by the rockets that will launch them into orbit—to form a larger instrument powerful enough to help physicists map spacetime right near the edge of a black hole.

Plans for Beyond Einstein also include three probes to complement the two large observatories. The Inflation Probe will detect gravitational waves unleashed in the moments right after the big bang, by measuring how they affected the microwave radiation that suffuses the universe. The Black Hole Finder Probe will search a large portion of the sky for hints of black holes, which, among other things, will help scientists pick targets for Constellation-X. The Dark Energy Probe—conceived of as a joint NASA/Department of Energy (DOE) mission—will survey the skies for supernovae. A large census of supernovae, which serve as cosmic yardsticks, will enable astrophysicists to home in on the properties of dark energy, the mysterious antigravity force that is causing the fabric of spacetime to expand faster and faster. "If you talk to anyone at any level at NASA

CREDIT: JPL/NASA

Published by AAAS

or DOE, they still seem very excited about it," says Saul Perlmutter, a supernova expert at Lawrence Berkeley National Laboratory in California who is working on one of the proposed designs for the dark energy mission. "I'm hoping you'll see all the important missions get a chance."

The dice game

When the Beyond Einstein project was launched in 2003, the auspices were good. Prioritization studies by physicists and astrophysicists gave LISA, Constellation-X, and some of the probes a very high rating. As a result, Congress backed the project and gave NASA the money it asked for to start Beyond Einstein on its way.

Then, a year ago, NASA slammed on the brakes. After Bush called for humans to return to the moon and eventually travel to Mars, White House and NASA managers diverted money from efforts like Beyond Einstein to get that program under way. NASA asked for only $40 million for the entire Beyond Einstein effort in 2005, delaying LISA and Constellation-X each by several years and indefinitely postponing the other missions (*Science*, 6 February 2004, p. 749). After receiving $25 million for LISA in 2004, NASA asked for only $19 million in 2005. Similarly, the proposed budget for Constellation-X dropped by nearly half, from $23.4 million to $12 million. And the agency slashed $1 million from the $10.5 million set aside in 2004 to start work on the other missions.

The decision to retrench stunned scientists. "I really hope the situation is going to change and NASA will take another look at their overall priorities," says Bender. "There's been a lot of work by the astrophysical community to determine their decadal prioritization, and the dropping of a substantial piece of that looks like a mistake."

Congress reluctantly complied with NASA's less enthusiastic plan in December, approving the Administration's request. A congressional aide says that many lawmakers were unhappy with NASA's decision to pull back but that the Beyond Einstein projects can't match the political clout of more mature projects. "They are not so entrenched yet, so that makes them vulnerable," the aide says. He predicts that without strong congressional pressure, Beyond Einstein funding will continue to be squeezed by NASA managers.

Worse may be yet to come. The agency still must allocate hundreds of millions of dollars in congressional earmarks, as well as space shuttle and Hubble Space Telescope costs, within its 2005 budget. And both NASA officials and outside scientists fear that younger efforts like Beyond Einstein will bear the brunt of those cuts, which likely will not be announced until late February, weeks after the 2006 White House budget request goes to Congress.

Stalled. Constellation-X's fleet of planned black hole–detecting x-ray satellites will remain an artist's conception until 2016 or beyond.

CREDIT: GSFC/NASA

One way around the financial squeeze might be to find allies with expertise and money. NASA and the European Space Agency (ESA) agreed last August to work together on the two separate missions that make up LISA, to the tune of $1 billion per agency. The first flight is a 2008 dress rehearsal known as LISA Pathfinder to test the advanced technologies to be used on the later LISA mission.

But already the tremendous complexity of the technologies has led to cost overruns and schedule delays. Two of the most difficult engineering challenges are to keep the innards of the LISA satellites on the correct path to within a nanometer or so and to ensure that disturbances such as solar photons and the spacecraft's own electrical systems don't affect the measurements. The key is a gravitational reference system, which will be NASA's main contribution to LISA Pathfinder. But cost increases on that system triggered a cancellation review last fall—which it survived—and another will take place in March. Bryant Cramer, Beyond Einstein program manager at Goddard Space Flight Center in Greenbelt, Maryland, says he is confident the system will survive the next scrutiny as well. In the meantime, technical challenges have postponed the launch of LISA Pathfinder from 2007 until the summer of 2008.

Once in space, LISA Pathfinder's results will immediately be fed into the design of LISA itself, in preparation for a 2013 launch of the three-satellite system. As a result, any delay to the first mission will ripple back, affecting not only LISA but possibly the other Beyond Einstein missions waiting in the queue.

After LISA comes Constellation-X. But plans for a 2013 or 2014 launch have already been abandoned in the wake of the cuts. Paul Geithner, Beyond Einstein program manager at NASA headquarters, says that the mission won't get off the ground before 2016, and other agency managers predict it will be several years later than that. As with LISA, funding trouble makes a partnership with ESA more attractive to NASA. The two agencies now are in negotiations to combine their efforts.

The three Einstein probes are next in line, although their fate is uncertain. "They haven't been dropped, merely put on hold until we understand the budget situation," says Cramer. Coordination between NASA and DOE on the Joint Dark Energy Mission, for example, continues despite the space agency's decision not to fund the effort. "They've got the money, and we don't," Cramer adds. Although legislative language in DOE's funding bill suggests DOE could take over the mission, "we still hope to partner with NASA," says Robin Staffin, director of DOE's high-energy physics program. A few advanced concept studies are under way, to the tune of $100,000 each, but "the real money doesn't kick in for some time," says Geithner.

Geithner and Cramer are nevertheless confident that the scientific promise of Beyond Einstein will ultimately carry the day. "Sensible people will see that these projects offer a whole new window into the universe," he says. "And Congress always resists" NASA's efforts to rob science to pay for space-flight missions, he notes. Cramer believes that LISA in particular has enough momentum to stay on track, because design work began in October with the approval of NASA science chief Al Diaz. "There is enough scientific cachet to do this mission, though it may be slower" than anticipated, he adds.

And Geithner contends that despite the budget competition, his program is here to stay. "It's going to happen," he insists. "Beyond Einstein is not going away."

–Charles Seife and Andrew Lawler

Published by AAAS

The Quantum Measurement Problem

A. J. Leggett

Despite the spectacular success of quantum mechanics (QM) over the last 80 years in explaining phenomena observed at the atomic and subatomic level, the conceptual status of the theory is still a topic of lively controversy. Most of the discussion centers around two famous paradoxes (or, as some would have it, pseudoparadoxes) associated, respectively, with the names of Einstein, Podolsky, and Rosen (EPR) and with Schrödinger's cat. In this Viewpoint, I will concentrate on the paradox of Schrödinger's cat or, as it is often known (to my mind somewhat misleadingly), the quantum measurement paradox.

Basically, the quantum measurement paradox is that most interpretations of QM at the microscopic level do not allow definite outcomes to be realized, whereas at the level of our human consciousness it seems a matter of direct experience that such outcomes occur (indeed, it seems so difficult to imagine what it would be like for the world to be otherwise that I suspect that Immanuel Kant, had he had occasion to consider the problem, would have classified our knowledge of this state of affairs as "synthetic a priori").

It is convenient to classify reactions to this problem into three broad classes, defined by the following three different views on the status of QM: (a) QM is the complete truth about the physical world, at all levels, and describes an external reality. (b) QM is the complete truth (in the sense that it will always give reliable predictions concerning the nature of experiments) but describes no external reality. (c) QM is not the complete truth about the world; at some level between that of the atom and that of human consciousness, other non–quantum mechanical principles intervene.

I briefly discuss each of these possibilities in turn [for a more extended discussion, see (*1*)]. Let's start with option (a). Consider the following two questions:

(1) In a typical situation involving an ensemble of microscopic entities (such as a Young's slits experiment with, for example, electrons or neutrons) in which the QM description of the ensemble is by a superposition of amplitudes corresponding to alternative microscopic possibilities A and B (e.g., "went through slit 1" and "went through slit 2"), is it the case that each individual member of the ensemble either definitely realizes alternative A or definitely realizes alternative B?

(2) In a (thought) experiment of the Schrödinger's cat type involving an ensemble of macroscopic objects (e.g., cats) for which the formal QM description of the ensemble of relevant "universes" is by a superposition of amplitudes corresponding to macroscopically distinct alternative states A and B (e.g., "cat alive" and "cat dead"), is it the case that each member (cat) of the ensemble either definitely realizes alternative A or definitely realizes alternative B (in the absence of inspection by a human agent)?

I believe that a large majority of the portion of the physics community that advocates option (a) would answer "no" to the first question and "yes" to the second (*2*).

The usual argument given in favor of these answers involves the phenomenon of decoherence: As a result of the latter phenomenon, it is impossible to see any effects of interference between (for example) the living and dead states of the cat, and it is argued that "therefore" one state or the other has been definitely realized, irrespective of whether we have or have not observed the particular cat in question.

As I have argued at greater length elsewhere, I believe this argument embodies a gross logical fallacy: It confuses the question of truth with the question of evidence. At the microscopic level, the adherents of view (a) felt (mostly) obliged to reject a realistic interpretation; the evidence they would cite against it is the well-known phenomenon of interference between possibilities A and B. By the time we get to the macroscopic level, the evidence has gone away, but the QM formalism is in no way changed; thus, its interpretation cannot have changed either.

To complete my argument at this point, it would be necessary to discuss also those interpretations of QM (such as the Bohm-de Broglie "hidden-variables" interpretation) that answer "yes" to both the above questions, and those (such as the Everett-Wheeler "many-universes" interpretation) that answer "no" to both. Because space is limited, I will just state my own view that both these interpretations amount to little more than verbal window dressing of the basic paradox, and thus that no interpretation of class (a) is viable (*3*).

I next turn more briefly to option (b). According to the adherents of this view, the whole formalism of QM amounts to nothing but a calculational recipe, designed in the last resort to predict the probabilities of various directly observed macroscopic outcomes ("this particular cat is dead/alive"), and the symbols occurring in it, such as the probability amplitudes, correspond to nothing in the "real world." The extreme operationalism implied in this view is often softened by the observation that under many conditions relevant to human existence, the experimental predictions of QM are "as if" the world had behaved classically; this argument is made most explicit in the "consistent-histories" (or "decoherent-histories") interpretation. However, that observation does not get around the fact that these conditions are not invariably fulfilled; in particular, it does not exclude a priori the possibility that we may some day be able not merely to generate quantum superpositions like that of Schrödinger's cat, but to observe the corresponding interference effects. Personally, if I could be sure that we will forever regard QM as the whole truth about the physical world, I think I should grit my teeth and plump for option (b).

Finally, what of option (c)? Indeed, there have been a number of concrete proposals to modify standard QM at some level intermediate between that of the atom and that of human consciousness, the currently best-developed one being probably that associated with the names of Ghirardi, Rimini, Weber, and Pearle. All of these proposals have in common the feature that at a sufficiently "macroscopic" level (the precise threshold depends on the specific proposal), the superpositions predicted by the formal extrapolation of the QM formalism do not occur; rather, some non-QM mechanism intervenes and guarantees the realization of a definite macroscopic outcome for each particular member of the ensemble in question. In principle, once the threshold for such realization is specified, it would be possible to test such theories unambiguously by comparing their predictions with those of standard QM; however, for the test to be definitive it is obviously necessary that QM continues to predict interference effects, i.e., that decoherence (which of course is a concept only meaningful within the QM formalism) has not washed them out. If one can indeed detect the characteristic QM interference effects at a given level of "macroscopicness," then it is a reason-

Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.

able inference [though not in itself a logically watertight one (*1*)] that no such mechanism of "realization" has come into play by that level.

Even a decade ago, considerable skepticism existed about the prospect of ever observing quantum superpositions involving more than a few "elementary" particles. However, in the last 5 years progress in this direction has been spectacular, ranging from traditional Young's slits experiments conducted with C_{70} molecules (~1300 "elementary" particles) to SQUID experiments in which the two superposed states involved $\sim 10^{10}$ electrons behaving differently (*1*). Thus, the experiments are beginning to impose nontrivial constraints on hypotheses of class (c). If in the future these constraints grow tighter and tighter, we may find that at the end of the day we have no alternative but to live with option (b).

References and Notes

1. A. J. Leggett, *J. Phys. Cond. Mat.* **14**, R415 (2002).
2. This belief is based on extensive canvassing of representative physics-colloquium audiences.
3. It is fair to warn any readers new to this topic that this conclusion is controversial in the extreme.

10.1126/science.1109541

VIEWPOINT

From Pedigree Cats to Fluffy-Bunnies

Jacob Dunningham, Alexander Rau, Keith Burnett*

We consider two distinct classes of quantum mechanical entanglement. The first "pedigree" class consists of delicate highly entangled states, which hold great potential for use in future quantum technologies. By focusing on Schrödinger cat states, we demonstrate not only the possibilities these states hold but also the difficulties they present. The second "fluffy-bunny" class is made up of robust states that arise naturally as a result of measurements and interactions between particles. This class of entanglement may be responsible for the classical-like world we see around us.

The nature of quantum superposition states and how we can "see" them in our classical world continues to fascinate scientists. In recent years, this fascination has led to a new awareness of the potential uses of these states in science and technology. Their nature opens the door to a whole range of new types of precision measurements. They also have important implications for what the classical world around us can look like. In this Viewpoint, we illustrate the nature of entanglement by focusing on two types of quantum states that we call "pedigree cats" and "fluffy-bunnies" (*1*). We want to explain why these states are so fascinating and why the pedigree cats are so difficult to breed and keep alive. They can be thought of as highly entangled, highly vulnerable, and easily killed off. The type of quantum entanglement that is breeding all around us and is responsible for the way we see the world is the wild fluffy-bunny kind.

The idea of a cat state first came about as a consequence of a famous thought experiment of Schrödinger in 1935 (*2*). In it, he imagined that a cat was placed in a box along with a radioactive sample arranged so that if a decay occurred, a toxic gas would be released and the cat killed. Quantum mechanics tells us that at any time the nucleus involved is in a superposition of the decayed and original state. Because the fate of the cat is perfectly correlated with the state of the nucleus undergoing decay, we are forced to conclude that the cat must also be in a superposition state, this time of being alive and dead. This result does not sit comfortably with our experience of the world around us—we would expect the cat to be either alive or dead but not both—and continues to fascinate and provoke discussion. Cat states have now come to refer to any quantum superposition of macroscopically distinct states. Here we call them pedigree cats to emphasize their prized but delicate nature.

Cat states are interesting not only for the questions they raise about quantum mechanics but also for their potential use in new quantum technologies. An important example of this is their use in pushing the limits of precision measurements. Because measurement is a physical process, we would expect the accuracy we can achieve in any measurement to be governed by the laws of physics. For quantum states, the very act of measuring changes the state and so affects subsequent results. This process is known as back-action. We will focus our discussion on interferometry, which is the basis for a wide range of precision measurements. Ultimately the precision that can be achieved in any measurement is subject to Heisenberg's uncertainty principle, which states that the uncertainty in any pair of conjugate variables obeys an inverse relation. The more accurately one variable is measured, the less accurately the other can be known. This leads to a fundamental limit to how accurately quantum phases can be measured that scales as $\Delta\phi \sim 1/N$, where N is the total number of particles involved. In practice, however, measurements are limited by more practical effects. Interferometry schemes, for example, usually use a stream of photons or atoms and are, therefore, normally limited by shot noise, where the measurement accuracy scales as $N^{-1/2}$. This conventional bound to measurement accuracy is a consequence both of the discrete nature of particles and of independent-particle statistics. The fundamental quantum limit (*3*, *4*) can be reached, however, by taking advantage of "cooperation" between the particles in entangled states. There are a number of proposals for how this might be achieved, and an excellent review of them is given by Giovannetti *et al*. (*5*). We will focus here on how entangled states, i.e., pedigree cats, open the door to this possibility.

If we were to split a single particle along the two paths of an interferometer, the state of the particle would be $|\Psi\rangle = (|1\rangle|0\rangle + e^{i\phi}|0\rangle|1\rangle)/\sqrt{2}$, where the first ket in each term represents the number of particles on one path and the second ket represents the number of particles on the other path. A particle on the second path acquires a phase shift ϕ relative to one on the first. Interferometry schemes generally use a stream of such single-particle states to make a measurement of ϕ. If instead we had a cat state of the form

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|N\rangle|0\rangle + |0\rangle|N\rangle) \qquad (1)$$

things would be quite different. The particles in this state are entangled because we cannot write the total state as a tensor product of the state of each of the particles. Another way of saying this is that if we know which way one of the particles goes, we know which way all of them go. This property makes these states very fragile—knowledge of the whereabouts of one particle blows the cover for all the others and destroys the superposition. However, this same property also makes the state very sensitive to phase shifts. In the case considered here, the phase shifts acquired

Clarendon Laboratory, Department of Physics, University of Oxford, Oxford OX1 3PU, UK.

*To whom correspondence should be addressed. E-mail: k.burnett1@physics.ox.ac.uk

able inference [though not in itself a logically watertight one (*1*)] that no such mechanism of "realization" has come into play by that level.

Even a decade ago, considerable skepticism existed about the prospect of ever observing quantum superpositions involving more than a few "elementary" particles. However, in the last 5 years progress in this direction has been spectacular, ranging from traditional Young's slits experiments conducted with C_{70} molecules (~1300 "elementary" particles) to SQUID experiments in which the two superposed states involved $\sim 10^{10}$ electrons behaving differently (*1*). Thus, the experiments are beginning to impose nontrivial constraints on hypotheses of class (c). If in the future these constraints grow tighter and tighter, we may find that at the end of the day we have no alternative but to live with option (b).

References and Notes

1. A. J. Leggett, *J. Phys. Cond. Mat.* **14**, R415 (2002).
2. This belief is based on extensive canvassing of representative physics-colloquium audiences.
3. It is fair to warn any readers new to this topic that this conclusion is controversial in the extreme.

10.1126/science.1109541

VIEWPOINT

From Pedigree Cats to Fluffy-Bunnies

Jacob Dunningham, Alexander Rau, Keith Burnett*

We consider two distinct classes of quantum mechanical entanglement. The first "pedigree" class consists of delicate highly entangled states, which hold great potential for use in future quantum technologies. By focusing on Schrödinger cat states, we demonstrate not only the possibilities these states hold but also the difficulties they present. The second "fluffy-bunny" class is made up of robust states that arise naturally as a result of measurements and interactions between particles. This class of entanglement may be responsible for the classical-like world we see around us.

The nature of quantum superposition states and how we can "see" them in our classical world continues to fascinate scientists. In recent years, this fascination has led to a new awareness of the potential uses of these states in science and technology. Their nature opens the door to a whole range of new types of precision measurements. They also have important implications for what the classical world around us can look like. In this Viewpoint, we illustrate the nature of entanglement by focusing on two types of quantum states that we call "pedigree cats" and "fluffy-bunnies" (*1*). We want to explain why these states are so fascinating and why the pedigree cats are so difficult to breed and keep alive. They can be thought of as highly entangled, highly vulnerable, and easily killed off. The type of quantum entanglement that is breeding all around us and is responsible for the way we see the world is the wild fluffy-bunny kind.

The idea of a cat state first came about as a consequence of a famous thought experiment of Schrödinger in 1935 (*2*). In it, he imagined that a cat was placed in a box along with a radioactive sample arranged so that if a decay occurred, a toxic gas would be released and the cat killed. Quantum mechanics tells us that at any time the nucleus involved is in a superposition of the decayed and original state. Because the fate of the cat is perfectly correlated with the state of the nucleus undergoing decay, we are forced to conclude that the cat must also be in a superposition state, this time of being alive and dead. This result does not sit comfortably with our experience of the world around us—we would expect the cat to be either alive or dead but not both—and continues to fascinate and provoke discussion. Cat states have now come to refer to any quantum superposition of macroscopically distinct states. Here we call them pedigree cats to emphasize their prized but delicate nature.

Cat states are interesting not only for the questions they raise about quantum mechanics but also for their potential use in new quantum technologies. An important example of this is their use in pushing the limits of precision measurements. Because measurement is a physical process, we would expect the accuracy we can achieve in any measurement to be governed by the laws of physics. For quantum states, the very act of measuring changes the state and so affects subsequent results. This process is known as back-action. We will focus our discussion on interferometry, which is the basis for a wide range of precision measurements. Ultimately the precision that can be achieved in any measurement is subject to Heisenberg's uncertainty principle, which states that the uncertainty in any pair of conjugate variables obeys an inverse relation. The more accurately one variable is measured, the less accurately the other can be known. This leads to a fundamental limit to how accurately quantum phases can be measured that scales as $\Delta\phi \sim 1/N$, where N is the total number of particles involved. In practice, however, measurements are limited by more practical effects. Interferometry schemes, for example, usually use a stream of photons or atoms and are, therefore, normally limited by shot noise, where the measurement accuracy scales as $N^{-1/2}$. This conventional bound to measurement accuracy is a consequence both of the discrete nature of particles and of independent-particle statistics. The fundamental quantum limit (*3*, *4*) can be reached, however, by taking advantage of "cooperation" between the particles in entangled states. There are a number of proposals for how this might be achieved, and an excellent review of them is given by Giovannetti *et al*. (*5*). We will focus here on how entangled states, i.e., pedigree cats, open the door to this possibility.

If we were to split a single particle along the two paths of an interferometer, the state of the particle would be $|\Psi\rangle = (|1\rangle|0\rangle + e^{i\phi}|0\rangle|1\rangle)/\sqrt{2}$, where the first ket in each term represents the number of particles on one path and the second ket represents the number of particles on the other path. A particle on the second path acquires a phase shift ϕ relative to one on the first. Interferometry schemes generally use a stream of such single-particle states to make a measurement of ϕ. If instead we had a cat state of the form

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|N\rangle|0\rangle + |0\rangle|N\rangle) \qquad (1)$$

things would be quite different. The particles in this state are entangled because we cannot write the total state as a tensor product of the state of each of the particles. Another way of saying this is that if we know which way one of the particles goes, we know which way all of them go. This property makes these states very fragile—knowledge of the whereabouts of one particle blows the cover for all the others and destroys the superposition. However, this same property also makes the state very sensitive to phase shifts. In the case considered here, the phase shifts acquired

Clarendon Laboratory, Department of Physics, University of Oxford, Oxford OX1 3PU, UK.

*To whom correspondence should be addressed. E-mail: k.burnett1@physics.ox.ac.uk

by particles in the second path combine to give the state

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|N\rangle|0\rangle + e^{iN\phi}|0\rangle|N\rangle). \quad (2)$$

This gives an N-fold enhancement in phase sensitivity over the single-particle case. However, in order to be even-handed, we should compare the performance of each scheme when the same number of atoms are used. Repeating the single-particle scheme N times gives a phase resolution that scales as $N^{-1/2}$. However, using the cat state gives a phase resolution that scales as N^{-1}, which is the fundamental quantum limit. This possibility of gaining a $\sqrt{N}$ improvement in measurement resolution is one of the reasons that cats are so highly prized. This shows how phase information can be exquisitely encoded on a quantum state, and various schemes have been put forward for achieving this. However, the important challenge is to devise techniques for reading out the information. This read-out should be regarded not as a mere detail but rather as a fundamental part of the physical process.

The key question is, how can we determine the phase between the two terms in Eq. 2? This is closely related to the question of how we can determine that this is indeed a coherent superposition rather than a classical mixture. The standard way to check for coherences is to perform an interference experiment. We can think of this in terms of Young's slits. In this experiment, two slits etched on an otherwise opaque barrier are illuminated with plane waves of light (or matter). On the other side of the barrier, a screen is set up to detect the positions of the particles that have passed through the slits (Fig. 1). It is well known that an interference pattern is seen and that this is due to each particle passing through both slits and interfering with itself when the two paths are recombined on the screen. If we were to include devices, D_1 and D_2, that record which slit each particle passes through, no interference would be seen. It is only by recombining the paths and ensuring that there is no information that betrays which path a particle took that we can confirm the superposition. The state of the photon at the slits can be considered to be catlike in the sense that the particle is in a superposition of two macroscopically distinct locations. The principle of Young's slits interference can therefore be applied as a general technique for seeing signatures of much larger cats.

Cat states have been created in the laboratory for atoms (*6*–*8*) and for buckminsterfullerene molecules ("bucky balls") (*9*, *10*) in two distinct locations, and interference experiments of the Young's slits type have confirmed that a superposition was created. This suggests a possible route for reading out the precision phases encoded on states of the form of (*3*): By simply overlapping the components on a screen and imaging an interference pattern, the position of the fringes should reveal the phase shift. However, this only works if the cat consists of a single object, e.g., a bucky ball. For the other types of cats, composed of collections of particles, this approach is less successful, as we will now see.

Suppose we had a cat state in a superposition of all N particles passing through one slit and all N particles passing through the other. How do we confirm that we have this superposition? If we had a multiparticle detector that detected all N particles at once, the situation would be equivalent to having one big object, and we would see interference in the multiparticle detections. If, however, we simply applied the Young's slits scheme and detected particles one at a time, i.e., the way we are usually forced to by available detectors, no interference would be seen.

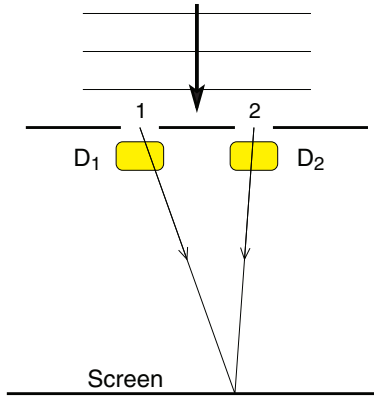


Fig. 1. Young's slits. Plane waves of particles are incident on a barrier with slits etched on it and are detected at a screen. Interference fringes are seen in the variation of intensity with position on the screen. If, however, detectors D_1 and D_2 record which slit each particle passes through, no interference is seen.

We can understand this because, although the detection on the screen cannot distinguish which path the particle took, we can in principle know which one it was. This is because, after the first detection, we could introduce detectors D_1 and D_2 to determine which path the second particle takes. Because all particles pass through the same slit, this also reveals the path that the first particle took. Interference is only seen if it is not possible, even in principle, to determine the particle's path (*11*) and, because it is possible here, no interference is seen. This same argument holds for subsequent measurements right up until the final one. The final detection will exhibit interference fringes because, after this detection, there is nothing left to reveal the path it took.

This result has been expressed elsewhere (*12*) in terms of correlation functions for the positions at which atoms are detected on the screen and, not surprisingly, only the Nth order correlation function reveals any difference between superposition states and mixtures.

In order to see cats, it seems that we not only need to recombine the elements of the superposition, but we also have to ensure that we wipe out any which-path information. This is the principle behind using probes to detect cats. We consider a state of the form of a cat entangled with a probe,

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|X_1\rangle|\uparrow\rangle + |X_2\rangle|\downarrow\rangle), \quad (3)$$

where $|X_1\rangle$ and $|X_2\rangle$ are states with the macroscopically distinct values X_1 and X_2 of some variable, e.g., position or momentum, and $|\rangle$ and $|\downarrow\rangle$ are the states of a single-particle probe, e.g., a photon that has different polarizations or passes through different slits. This has the same form as Schrödinger's original idea of a cat being dead and alive entangled with a nucleus that is decayed and not. If we now perform an interference experiment on the probe state, the probability of detecting a particle corresponding to a phase shift ϕ between the paths is $P_\phi = \frac{1}{2}[1 + \Re(e^{i\phi}\langle X_1|X_2\rangle)]$. If the macroscopic states are orthogonal, i.e., $\langle X_1|X_2\rangle = 0$, then there is no interference. If, however, we could operate on state 3 in such a way as to make the macroscopic states indistinguishable, i.e., $\langle X_1|X_2\rangle = 1$, we get $P_\phi = 1 + \cos\phi$, and fringes are visible. This is the idea behind a number of cat schemes, including an experimental study of how superpositions of coherent states of light are affected by loss (*13*) and a theoretical proposal for how cats may be created in the motion of micromirrors (*14*).

In essence, this read-out scheme scarcely differs from the Young's slits scheme, because it involves single-particle interference with no which-path information. The state we are left with, $|X_0\rangle$, is no longer a cat, and our only evidence for the cat's former existence is the faintest murmur of a death cry. It would be nice to have a scheme that provides evidence for the cat without destroying it. One possibility for achieving this is to only "partially recombine" the components of the cat. By this we mean that we perform some operation on Eq. 3 to obtain a state of the form,

$$|\Psi\rangle = \frac{1}{\sqrt{2}}[(a|X_1\rangle + b|X_0\rangle)|\uparrow\rangle + (a|X_2\rangle + b|X_0\rangle)|\downarrow\rangle], \quad (4)$$

where $|a|^2 + |b|^2 = 1$ and, for weak recombination, $|b| \ll |a|$. In this case, we have partially wiped out the which-path information and, if we performed an interference experiment with the probe state, we

would see fringes with reduced visibility $|b|^2$. The key point, however, is that, after detecting the probe, the state we are left with is $|\Psi\rangle = a(|X_1\rangle + |X_2\rangle)/\sqrt{2} + \sqrt{2}b|X_0\rangle$, i.e., $|\Psi\rangle \approx (|X_1\rangle + |X_2\rangle)/\sqrt{2}$, and so the macroscopic superposition is only slightly affected. In essence, so long as we can identify fringes with small visibilities, it is possible to find signatures of cats while only delicately changing them. This further emphasizes the important role that precision measurements play in quantum physics.

Although we have found a way to avoid destroying the cats, we are still drawn to the view that recombination (whether complete or partial) is necessary to see them (*15*). If this is true, there are some interesting philosophical consequences. In particular, if no operator exists that can recombine a certain type of cat, then we can never see that cat. For example, in Schrödinger's original experiment, when we open the box, what we see may indeed be a superposition of the cat being alive and dead. However, because there is no "Lazarus operator" that turns dead into alive, there can be no way of distinguishing this from a mixture. We are free to choose the interpretation we wish because there are no physical consequences of picking one in favor of the other. It is much more appealing to our instincts of the world around us to think of it as a mixture, and so this is how we generally interpret it.

We have seen that cat states are very useful but temperamental. They prove difficult to observe, and their form means they are fragile with superpositions that are destroyed by the slightest dissipation (*16*–*19*). There are also other highly prized pedigree species that form the general class of highly entangled states and, like cats, are well worth studying because of the possibilities they afford in measurement schemes (*5*) and other quantum technologies. Partly because of their fragility and partly because they prove difficult to observe, these pedigree states are not familiar to our everyday experience of the world. What we see instead is a whole different class of states that we call fluffy-bunnies. These are the robust entanglements that arise as a result of measurements and interactions between particles.

The idea of fluffy-bunnies developed from theoretical work undertaken to study the interference fringes that are seen when two Bose-Einstein condensates spatially overlap (*20*). If these condensates are both initially in number states, they contain no quantum phase information, and we would not expect any interference. This is seen from $\Delta N \Delta\phi \sim 1$: If $\Delta N \to 0$, $\Delta\phi \to \infty$. However, a careful analysis of the measurement process reveals the surprising result that interference fringes are in fact seen.

This can be understood because each detection of an atom entangles the two condensates due to the fact that we do not know which condensate the atom came from. This entanglement induces a relative phase between the condensates, which serves to reinforce the probability of certain positions for the next atom detected, and the process continues by a feedback mechanism. An interference pattern builds up on the screen, and the condensates develop a well-defined relative phase. The condensates are now in a fluffy-bunny state: They have acquired a relative observable, which is robust, i.e., it hardly changes with subsequent detections and so is, in a sense, classical. The width of the relative phase distribution scales as $N^{-1/2}$ for the number of atoms detected and so gives us an understanding of the conventional measurement bound for nonpedigree states.

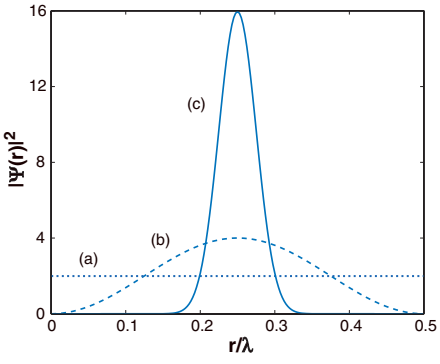


Fig. 2. Relative position probability distribution for an initially delocalized pair of particles after (a) 0, (b) 1, and (c) 20 photons have been scattered from them and detected. The relative position of the particles is plotted in units of the wavelength, λ, of the photons. We see that the relative localization becomes better defined as more detections are made.

These fluffy-bunnies are fascinating and reveal how classical-like variables can emerge from quantum systems. In the case considered here, only relative phases exist; the absolute phase of each condensate remains undefined throughout the process. Furthermore, these phases are transitive so that if we have a collection of condensates and know the phase of each of them relative to any one of them, then we know the relative phase of any pair of condensates. This means that the classical nature of fluffy-bunny entanglements enables us to define a consistent phase standard for Bose-Einstein condensates (*21*). This concept of transitivity is ingrained in our classical perception of the world but is not obvious in quantum mechanics, where measurements generally change the system.

Fluffy-bunnies can be applied equally well to any pair of conjugate variables. For example, they can be used to understand how objects localize in position space—an issue central to the boundary between quantum and classical physics. An analysis has been carried out for light scattering from a pair of particles that are intially smeared out over space, e.g., the particles are each initially in momentum eigenstates (*22*). A sequence of measurements of the scattered photons entangles the two particles, leading to robust semiclassical states of well-defined relative position. This can be seen by considering a system of two particles in momentum eigenstates with relative momentum p, i.e., $|\Psi\rangle = |p\rangle$. After scattering a photon from these particles and detecting the angle at which it is scattered, the state of the particles is

$$|\Psi\rangle = \frac{1}{\sqrt{2}}\left[|p + \frac{\Delta}{2}\rangle + e^{i\phi}|p - \frac{\Delta}{2}\rangle\right], \quad (5)$$

where Δ is the momentum kick imparted by the photon and ϕ is a phase shift that depends on the angle at which the photon is detected. We see that the measurement has broadened the relative momentum distribution of the particles. Because the system remains in a pure state, Heisenberg's uncertainty relation tells us that this must be accompanied by a reduction in the relative position distribution. This process continues by feedback, and Fig. 2 shows the relative position distribution of the particles after (a) 0, (b) 1, and (c) 20 scattering events. The initially delocalized particles become progressively more localized as more measurements are made, and these relative positions are robust because subsequent measurements do not change their mean value. In this sense, we can think of them as classical variables and can see why fluffy-bunnies describe a world that is much more familiar to our everyday experience. As was the case for the phase of condensates, these localizations are transitive (*23*) and so form a consistent coordinate space. The absolute positions of the particles remain undefined throughout the procedure, which draws us toward the interesting conclusion that relative observables are fundamental in our world.

The pedigree and fluffy-bunny classes of states offer a fascinating insight into quantum physics. Delicate pedigree entanglements give us a fleeting glimpse of the quantum world and hint at the phenomenal potential for new technologies it contains. Their fragile nature, however, means that they do not describe the world we see every day. This role is left to fluffy-bunnies—the wild and hardy states that breed all around us. There is a lot more to be understood about these different classes of states, not least the profound implications that quantum measurement has on the way we see the world and the rich potential it keeps largely hidden from our view.

References and Notes

1. A "fluffy-bunny" is a cheap, manufactured toy given as a prize in British fairgrounds.
2. E. Schrödinger, *Naturwissenschaften* **23**, 807 (1935).
3. J. J. Bollinger, W. M. Itano, D. J. Wineland, D. J. Heinzen, *Phys. Rev. A* **54**, R4649 (1996).
4. Z. Y. Ou, *Phys. Rev. A* **55**, 2598 (1997).
5. V. Giovannetti, S. Lloyd, L. Maccone, *Science* **306**, 1330 (2004).
6. O. Carnal, J. Mlynek, *Phys. Rev. Lett.* **66**, 2689 (1991).
7. D. W. Keith, C. R. Ekstrom, Q. A. Turchette, D. E. Pritchard, *Phys. Rev. Lett.* **66**, 2693 (1991).
8. M. Kasevich, S. Chu, *Phys. Rev. Lett.* **67**, 181 (1991).
9. M. Arndt *et al.*, *Nature* **401**, 680 (1999).
10. L. Hackermüller, K. Hornberger, B. Brezger, A. Zeilinger, M. Arndt, *Nature* **427**, 711 (2004).
11. S. M. Tan, D. F. Walls, *Phys. Rev. A* **47**, 4663 (1993).
12. R. Bach, K. Rzążewski, *Phys. Rev. Lett.* **92**, 200401 (2004).
13. M. Brune *et al.*, *Phys. Rev. Lett.* **77**, 4887 (1996).
14. W. Marshall, C. Simon, R. Penrose, D. Bouwmeester, *Phys. Rev. Lett.* **91**, 130401 (2003).
15. Another way to find evidence for a cat, not discussed here, is to disentangle the particles, but this also amounts to recombination and destroys the cat.
16. E. Joos, H. D. Zeh, *Z. Phys. B* **59**, 223 (1985).
17. W. H. Zurek, *Phys. Today* **44**, 36 (1991).
18. G. C. Ghirardi, A. Rimini, T. Weber, *Phys. Rev. D* **34**, 470 (1986).
19. W. H. Zurek, *Rev. Mod. Phys.* **75**, 715 (2003).
20. J. Javanainen, S. M. Yoo, *Phys. Rev. Lett.* **76**, 161 (1996).
21. J. A. Dunningham, K. Burnett, *Phys. Rev. Lett.* **82**, 3729 (1999).
22. A. V. Rau, J. A. Dunningham, K. Burnett, *Science* **301**, 1081 (2003).
23. J. A. Dunningham, A. V. Rau, K. Burnett, *J. Mod. Opt.* **51**, 2323 (2004).

24. This work was supported by the UK Engineering and Physical Sciences Research Council and the Royal Society and Wolfson Foundation.

10.1126/science.1109545

REVIEW

Time and the Quantum: Erasing the Past and Impacting the Future

Yakir Aharonov[1,2] and M. Suhail Zubairy[3]*

The quantum eraser effect of Scully and Drühl dramatically underscores the difference between our classical conceptions of time and how quantum processes can unfold in time. Such eyebrow-raising features of time in quantum mechanics have been labeled "the fallacy of delayed choice and quantum eraser" on the one hand and described "as one of the most intriguing effects in quantum mechanics" on the other. In the present paper, we discuss how the availability or erasure of information generated in the past can affect how we interpret data in the present. The quantum eraser concept has been studied and extended in many different experiments and scenarios, for example, the entanglement quantum eraser, the kaon quantum eraser, and the use of quantum eraser entanglement to improve microscopic resolution.

The "classical" notion of time was summed up by Newton: "…absolute and mathematical time, of itself, and from its own nature, flows equally without relation to anything external." In the present article, we go beyond our classical experience by presenting counterintuitive features of time as it evolves in certain experiments in quantum mechanics. To illustrate this point, an excellent example is the delayed-choice quantum eraser, proposed by Marlan O. Scully and Kai Drühl (*1*), which was described as an idea that "shook the physics community" when it was first published in 1982 (*2*). They analyzed a photon correlation experiment designed to probe the extent to which information accessible to an observer and its erasure affects measured results. The Scully-Drühl quantum eraser idea as it was described in *Newsweek* tells the story well (*3*), and Fig. 1 is an adaptation of their account of this fascinating effect.

In his book *The Fabric of the Cosmos* (*4*), Brian Greene sums up beautifully the counterintuitive outcome of the experimental realizations of the Scully-Drühl quantum eraser (p. 149):

> *These experiments are a magnificent affront to our conventional notions of space and time. Something that takes place long after and far away from something else nevertheless is vital to our description of that something else. By any classical-common sense-reckoning, that's, well, crazy. Of course, that's the point: classical reckoning is the wrong kind of reckoning to use in a quantum universe …. For a few days after I learned of these experiments, I remember feeling elated. I felt I'd been given a glimpse into a veiled side of reality. Common experience—mundane, ordinary, day-to-day activities—suddenly seemed part of a classical charade, hiding the true nature of our quantum world. The world of the everyday suddenly seemed nothing but an inverted magic act, lulling its audience into believing in the usual, familiar conceptions of space and time, while the astonishing truth of quantum reality lay carefully guarded by nature's sleights of hand.*

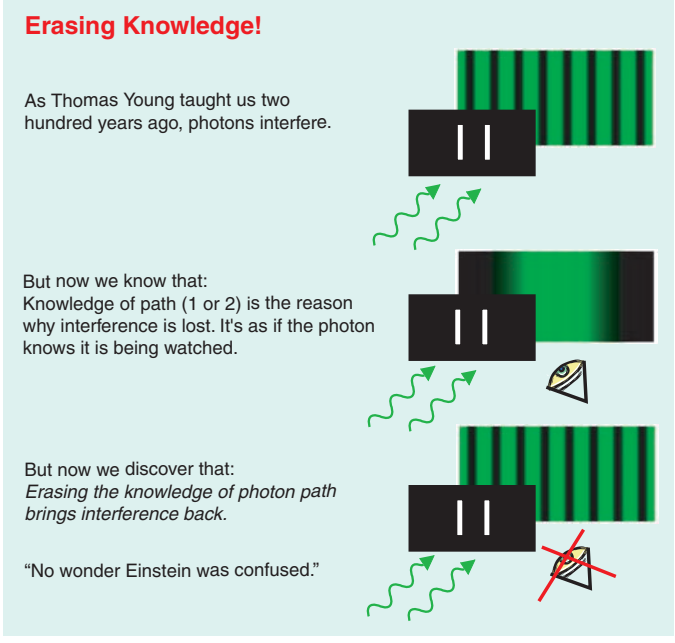


Fig. 1. Schematics for the Young's double-slit experiment. The which-path information wipes out the interference pattern. The interference pattern can be restored by erasing the which-path information.

[1]School of Physics and Astronomy, Tel Aviv University, Tel Aviv 69978, Israel. [2]Department of Physics, University of South Carolina, Columbia, SC 29208, USA. [3]Institute for Quantum Studies and Department of Physics, Texas A&M University, College Station, TX 77843, USA.

*To whom correspondence should be addressed. E-mail: zubairy@physics.tamu.edu

References and Notes

1. A "fluffy-bunny" is a cheap, manufactured toy given as a prize in British fairgrounds.
2. E. Schrödinger, *Naturwissenschaften* **23**, 807 (1935).
3. J. J. Bollinger, W. M. Itano, D. J. Wineland, D. J. Heinzen, *Phys. Rev. A* **54**, R4649 (1996).
4. Z. Y. Ou, *Phys. Rev. A* **55**, 2598 (1997).
5. V. Giovannetti, S. Lloyd, L. Maccone, *Science* **306**, 1330 (2004).
6. O. Carnal, J. Mlynek, *Phys. Rev. Lett.* **66**, 2689 (1991).
7. D. W. Keith, C. R. Ekstrom, Q. A. Turchette, D. E. Pritchard, *Phys. Rev. Lett.* **66**, 2693 (1991).
8. M. Kasevich, S. Chu, *Phys. Rev. Lett.* **67**, 181 (1991).
9. M. Arndt *et al.*, *Nature* **401**, 680 (1999).
10. L. Hackermüller, K. Hornberger, B. Brezger, A. Zeilinger, M. Arndt, *Nature* **427**, 711 (2004).
11. S. M. Tan, D. F. Walls, *Phys. Rev. A* **47**, 4663 (1993).
12. R. Bach, K. Rzążewski, *Phys. Rev. Lett.* **92**, 200401 (2004).
13. M. Brune *et al.*, *Phys. Rev. Lett.* **77**, 4887 (1996).
14. W. Marshall, C. Simon, R. Penrose, D. Bouwmeester, *Phys. Rev. Lett.* **91**, 130401 (2003).
15. Another way to find evidence for a cat, not discussed here, is to disentangle the particles, but this also amounts to recombination and destroys the cat.
16. E. Joos, H. D. Zeh, *Z. Phys. B* **59**, 223 (1985).
17. W. H. Zurek, *Phys. Today* **44**, 36 (1991).
18. G. C. Ghirardi, A. Rimini, T. Weber, *Phys. Rev. D* **34**, 470 (1986).
19. W. H. Zurek, *Rev. Mod. Phys.* **75**, 715 (2003).
20. J. Javanainen, S. M. Yoo, *Phys. Rev. Lett.* **76**, 161 (1996).
21. J. A. Dunningham, K. Burnett, *Phys. Rev. Lett.* **82**, 3729 (1999).
22. A. V. Rau, J. A. Dunningham, K. Burnett, *Science* **301**, 1081 (2003).
23. J. A. Dunningham, A. V. Rau, K. Burnett, *J. Mod. Opt.* **51**, 2323 (2004).

24. This work was supported by the UK Engineering and Physical Sciences Research Council and the Royal Society and Wolfson Foundation.

10.1126/science.1109545

REVIEW

Time and the Quantum: Erasing the Past and Impacting the Future

Yakir Aharonov[1,2] and M. Suhail Zubairy[3]*

The quantum eraser effect of Scully and Drühl dramatically underscores the difference between our classical conceptions of time and how quantum processes can unfold in time. Such eyebrow-raising features of time in quantum mechanics have been labeled "the fallacy of delayed choice and quantum eraser" on the one hand and described "as one of the most intriguing effects in quantum mechanics" on the other. In the present paper, we discuss how the availability or erasure of information generated in the past can affect how we interpret data in the present. The quantum eraser concept has been studied and extended in many different experiments and scenarios, for example, the entanglement quantum eraser, the kaon quantum eraser, and the use of quantum eraser entanglement to improve microscopic resolution.

The "classical" notion of time was summed up by Newton: "…absolute and mathematical time, of itself, and from its own nature, flows equally without relation to anything external." In the present article, we go beyond our classical experience by presenting counterintuitive features of time as it evolves in certain experiments in quantum mechanics. To illustrate this point, an excellent example is the delayed-choice quantum eraser, proposed by Marlan O. Scully and Kai Drühl (*1*), which was described as an idea that "shook the physics community" when it was first published in 1982 (*2*). They analyzed a photon correlation experiment designed to probe the extent to which information accessible to an observer and its erasure affects measured results. The Scully-Drühl quantum eraser idea as it was described in *Newsweek* tells the story well (*3*), and Fig. 1 is an adaptation of their account of this fascinating effect.

In his book *The Fabric of the Cosmos* (*4*), Brian Greene sums up beautifully the counterintuitive outcome of the experimental realizations of the Scully-Drühl quantum eraser (p. 149):

> *These experiments are a magnificent affront to our conventional notions of space and time. Something that takes place long after and far away from something else nevertheless is vital to our description of that something else. By any classical-common sense-reckoning, that's, well, crazy. Of course, that's the point: classical reckoning is the wrong kind of reckoning to use in a quantum universe …. For a few days after I learned of these experiments, I remember feeling elated. I felt I'd been given a glimpse into a veiled side of reality. Common experience—mundane, ordinary, day-to-day activities—suddenly seemed part of a classical charade, hiding the true nature of our quantum world. The world of the everyday suddenly seemed nothing but an inverted magic act, lulling its audience into believing in the usual, familiar conceptions of space and time, while the astonishing truth of quantum reality lay carefully guarded by nature's sleights of hand.*

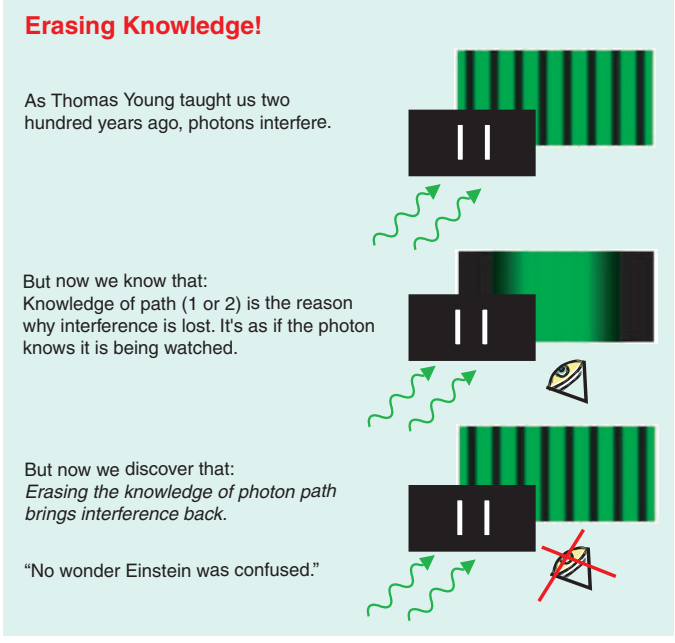


Fig. 1. Schematics for the Young's double-slit experiment. The which-path information wipes out the interference pattern. The interference pattern can be restored by erasing the which-path information.

[1]School of Physics and Astronomy, Tel Aviv University, Tel Aviv 69978, Israel. [2]Department of Physics, University of South Carolina, Columbia, SC 29208, USA. [3]Institute for Quantum Studies and Department of Physics, Texas A&M University, College Station, TX 77843, USA.

*To whom correspondence should be addressed. E-mail: zubairy@physics.tamu.edu

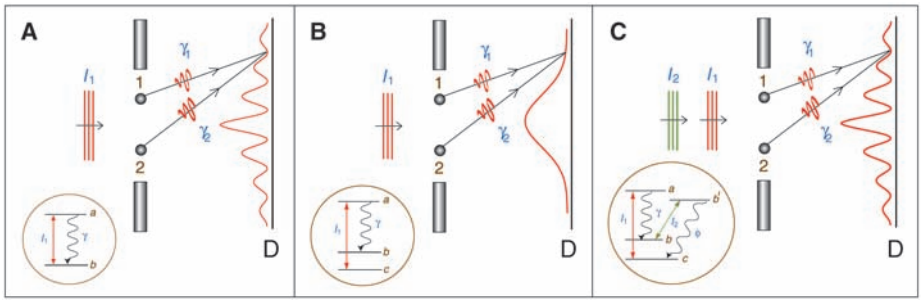


Fig. 2. Here, we consider three possible configurations of atoms that are placed at sites 1 and 2. In (**A**) we consider a two-level atom initially in the state b. The incident pulse l_1 excites one of the two atoms to state a from where it decays to state b, emitting a γ photon. In (**B**) the atom initially in the ground state c is excited by the pulse l_1 to state a from where it decays to state b. In (**C**) a fourth level is added. A pulse l_2 excites the atom to state b' after the atom has decayed to state b. The atom in the state b' emits a ϕ photon and ends up in state c.

Quantum Eraser Basics

We now present a simple description of the quantum eraser that brings out the counterintuitive aspects related to time in the quantum mechanical domain. We consider the scattering of light from two atoms located at sites 1 and 2 on the screen D (Fig. 2) and analyze three different cases:

1) Resonant light impinges from the left on two-level atoms (Fig. 2A) located at sites 1 and 2. An atom excited to level a emits a γ photon. There are two possibilities for the atom, either it remains in the ground state b or it can get excited to the state a by the incident light and emit a γ photon. We look at the interference of these photons at the screen. Because both atoms are finally in the state b after the emission of photons, it is not possible to determine which atom contributed the γ photon. A large number of such experiments are carried out; i.e., any one photon will yield one count on the screen, and it takes many such photon events to build up a pattern. The resulting distribution of the detected photons exhibits an interference pattern (Fig. 2A). This is an analog of the usual Young's double-slit experiment. Instead of the usual light beams through two pin holes, we have considered scattered light from two atoms. The key to the appearance of the interference is the lack of which-path information for the photons.

2) In the case where the atoms have three levels (Fig. 2B), the drive field excites the atoms from the ground state c to the excited state a. The atom in state a can then emit a γ photon and end up in state b. Here, the photon detected on the screen leaves behind which-path information; that is, the atom responsible for contributing the γ photon is in level b, whereas the other atom remains in level c. Thus, a measurement of the internal states of the atoms provides us the which-path information and no interference is observed. That is, the state of the atom acts as an observer state. The precise mathematical description of photons γ_1 and γ_2 is the same in cases a and b. It is only the presence of the passive observer state that kills the interference.

There is an interesting connection to be made here with a statement of Richard Feynman. In his wonderful lectures on quantum mechanics for Caltech undergraduates (*5*), he discusses the problem of such observations rubbing out interference. He says (p. 9)

> *If an apparatus is capable of determining which hole the [photon] goes through it cannot be so delicate that it does not disturb the pattern in an essential way. No one has ever found or even thought of a way around the uncertainty principle. So we must assume that it describes a basic characteristic of nature.*

However, the loss of coherence in the present scheme does not invoke the uncertainty principle. In later work, Englert, Schwinger, Scully, and Walther came up with other such examples and in this sense have "thought of a way around the uncertainty principle" in this regard. We discuss this below.

The question, however, is whether we can erase the which-path information stored in the atom(s) and thus regain interference. If the loss of interference was caused by some kind of noise or uncertainty due to quantum fluctuations, the answer would be no. We now show that this is not the case, and the interference can be recovered. The question then is whether it is possible to wipe out the which-path information and recover the interference.

3) As shown in Fig. 2C, this can possibly be done by driving the atom by another field that takes the atom from level b to b' and, after an emission of a ϕ photon at the $b' - c$ transition, ends up in level c. Now the final state of both the atoms is c, and a measurement of internal states cannot provide us the which-path information. It would therefore seem that the interference fringes will be restored, but a careful analysis indicates that the which-path information is still available through the ϕ photon. A measurement on the ϕ photon can tell us which atom contributed the γ photon. Can we erase the which-path information contained in the ϕ photon and recover the interference fringes? Scully and Drühl considered an ingenious device based on an electrooptic shutter that can absorb the ϕ photon in such a way that the which-path information is erased (*1*). For the purpose of illustration, we consider a different and somewhat simplified version of such an eraser. A slightly modified version of such an eraser using a parametric process involving nonlinear crystal (instead of single atoms) was experimentally realized by Shih and co-workers in 2000 (*6*), which served as the motivation for Greene's presentation in (*4*).

However, before we proceed with discussions of Shih's experiment we should note that the erasure idea stirred up considerable controversy. Perhaps the best example is the well-written article by Mohrhoff (*7*). In the abstract, which we

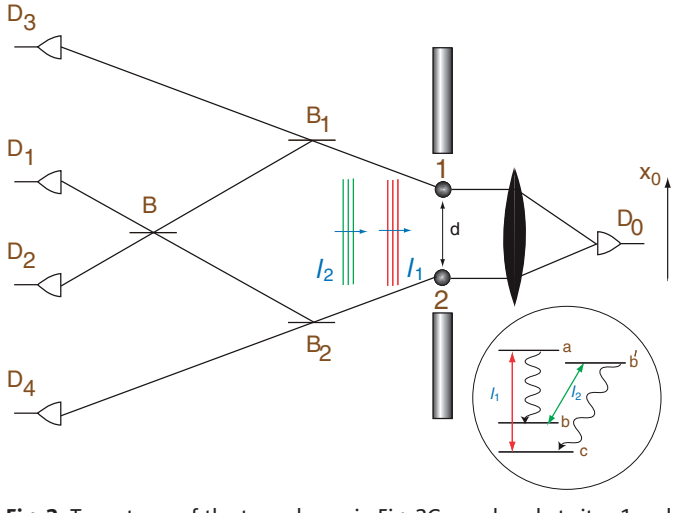


Fig. 3. Two atoms of the type shown in Fig. 2C are placed at sites 1 and 2. These atoms are excited by pulses l_1 and l_2 as in Fig. 2C such that one of the atoms emits γ and ϕ photons. We consider those events where the γ photon proceeds to the right and the ϕ photon to the left. The γ photon is collected by the detector D_0, whereas the ϕ photon is detected by D_1, D_2, D_3, or D_4 after passing through the optical setup consisting of the 50/50 beam splitters B_1, B_2, and B.

have adapted to fit the present example, he says (p. 1468)

> *a two-slit experiment…appears to permit experimenters to choose even after each photon has made its mark on the screen, whether the photon has passed through a particular slit or has, in some sense, passed through both of them. Through a misleading wording the authors even appear to endorse this interpretation.*

In a later paper, however, the author retracts this statement (*8*).

In fact, many people had a similar mind set, and it is only by carefully considering and analyzing several experiments (real and *gedanken*) that the issue is made clear.

We now turn to the particularly clear treatment of Shih and coworkers as depicted in Fig. 3. We again consider two atoms of the type shown in Fig. 2C located at sites 1 and 2. A pair of photons γ and φ are emitted either by the atom located at 1 or by the atom located at 2. The γ photon, as before, proceeds to the screen on the right and is detected by a detector on screen D at a location x_0. A repeat of this experiment yields an essentially random distribution of photons on the screen.

What about the appearance and disappearance of interference fringes discussed above? For this purpose, we look at the φ photon that proceeds to the left. We consider only those instances where the φ photon scattered from the atom located at 1 proceeds to the beam splitter B_1 and the φ photon scattered from the atom located at 2 proceeds to B_2. At either of these 50/50 beam splitters, the φ photon has a 50% probability of proceeding to detectors D_3 (for photon scattered from 1) and to D_4 (for photon scattered from 2). On the other hand, there is also a 50% probability that the photon will be reflected from the respective beam splitter and proceed to another 50/50 beam splitter, B. For these photons, there is an equal probability of being detected at detectors D_1 and D_2.

If the φ photon is detected at the detector D_3, it has necessarily come from the atom located at 1 and could not have come from the atom located at 2. Similarly, detection at D_4 means that the φ photon came from the atom located at 2. For such events, we can also conclude that the corresponding γ photon was also scattered from the same atom. That is, we have "which-way" information if detectors D_3 or D_4 register a count.

Returning to the quantum erasure protocol, if the φ photon is detected at D_1, there is an equal probability that it may have come from the atom located at 1, following the path $1B_1BD_1$, or it may have come from the atom located at 2, following the path $2B_2BD_1$. Thus, we have erased the information about which atom scattered the φ photon, and there is no which-path information available for the corresponding γ photon. The same can be said about the φ photon detected at D_2. The difference between counts in D_1 and D_2 is a phase shift such that a click at D_1 gives the fringes corresponding to $\gamma_1 + \gamma_2$, whereas a click at D_2 correlates with $\gamma_1 - \gamma_2$.

After this experiment is done a large number of times, we shall have roughly 25% of φ photons detected each at D_1, D_2, D_3, and D_4 because of the 50/50 nature of our beam splitters. The corresponding spatial distribution of γ photons will be, as mentioned above, completely random. Next we do a sorting process. We separate out all the events where the φ photons are detected at D_1, D_2, D_3, and D_4. For these four groups of events, we locate the positions of the detected γ photons on the screen D.

The key result is that, for the events corresponding to the detection of φ photons at detectors D_3 and D_4, the pattern obtained by the γ photons on the screen D is the same as we would expect if these photons had scattered from atoms at sites 1 and 2, respectively. That is, there are no interference fringes, as would be expected when we have which-path information available. On the contrary, we obtain conjugate (π phase shifted) interference fringes for those events where the φ photons are detected at D_1 and D_2. For this set of data, there is no which-path information available for the corresponding γ photons.

Suppose we place the φ photon detectors far away. Then the future measurements on these photons influence the way we think about the γ photons measured today (or yesterday!). For example, we can conclude that γ photons whose φ partners were successfully used to ascertain which-path information can be described as having (in the past) originated from site 1 or site 2. We can also conclude that γ photons whose φ partners had their which-path information erased cannot be described as having (in the past) originated from site 1 or site 2 but must be described, in the same sense, as having come from both sites. The future helps shape the story we tell of the past.

Here again the eloquent and insightful Brian Greene says it well (p. 197):

> *Notice, too, perhaps the most dazzling result of all: the three additional beam*

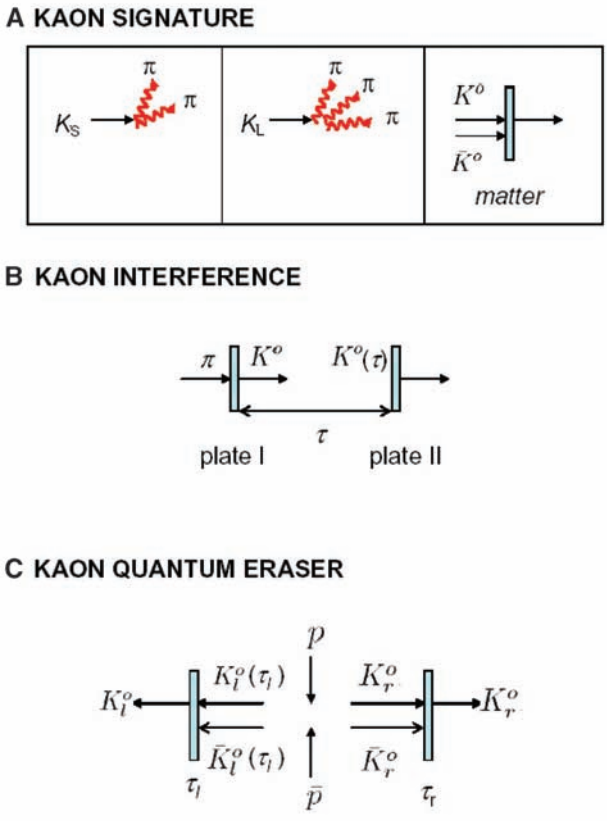


Fig. 4. (**A**) The four kaons K_S, K_L, K^0, and $\bar{K}^0$ have characteristic signatures; (the short-lived) kaon K_S decays into two π particles, whereas (the long-lived) kaon K_L decays into three π particles; the K^0 kaon (strangeness +1) mostly passes through matter, but the $\bar{K}^0$ (strangeness −1) interacts much more strongly with matter (nuclei) and is stopped. (**B**) The K^0 and $\bar{K}^0$ states are superpositions of K_S and K_L, i.e., $|K^0\rangle = (|K_S\rangle + |K_L\rangle)/\sqrt{2}$ and $|\bar{K}^0\rangle = (|K_S\rangle - |K_L\rangle)/\sqrt{2}$. Now K_S and K_L have masses m_S and m_L so that $|\bar{K}^0(\tau)\rangle = (e^{-im_S\tau}|K_S\rangle + e^{-im_L\tau}|K_L\rangle)/\sqrt{2}$. Thus, if we produce K^0 particles in plate I and they propagate for a time τ to plate II then the probability for passage through plate II is $|\langle K^0|K^0(\tau)\rangle|^2$ which shows oscillations in time. (**C**) A kaon quantum eraser may be realized by noting that $p\,\bar{p}$ collisions generate the entangled states moving to the right (r) and left (l) which can be written in terms of which-way (K_S, K_L) or which-wave (K^0, $\bar{K}^0$). Quantum erasing is achieved by the left-moving kaon as the measured kaon (which will or will not show oscillations), and the right tag or ancilla kaon will serve to select the which-wave ensemble (K^0, $\bar{K}^0$) if we put in plate II and measure K_r^0. However, if we do not put in the second plate then we must describe the physics by the which-way subensemble. Thus, the entangled kaon state can be used to demonstrate quantum erasure by subensemble selection just as in the original photon case. However, if K_S or K_L propagates from I to II, the state of the kaon just before it enters II is $|K_S(\tau)\rangle = e^{-im_S\tau}|K_S\rangle$ and $|\langle K^0|K_S(\tau)\rangle|^2 = 1/2$ with a similar result for K_L. In this sense, K_S and K_L are "which-way" (short or long lived) states like photons going through slit 1 or 2, i.e., do not show oscillations. K^0 and $\bar{K}^0$, however, do show oscillation behavior and in this sense may be called "which-wave."

splitters and the four idler-photon detectors can only be on the other side of the laboratory or even on the other side of the universe, since nothing in our discussion depended at all on whether they receive a given idler photon before or after its signal photon partner has hit the screen. Imagine, then, that these devices are all far away, say ten light-years away, to be definite, and think about what this entails. You perform the experiment in fig 7.5b today, recording–one after another–the impact locations of a huge number of signal photons and you observe that they show no sign of interference. If someone asks you to explain the data, you might be tempted to say that because of the idler photons, which path information is available and hence each signal photon definitely went along either the left or the right path, eliminating any possibility of interference. But, as above, this would be a hasty conclusion about what happened; it would be a thoroughly premature description of the past.

For the mathematically inclined reader we include a brief discussion (*9*) which sheds light on the physics using the language of modern quantum mechanics.

The Micromaser Which-Path Detector and Quantum Eraser

The Scully-Drühl quantum eraser was perhaps the earliest example of quantum entanglement interferometry and stimulated many experiments. However, another form of the quantum eraser based on cavity quantum electrodynamics and the micromaser has also stimulated debate as well as new experiments and calculations. In particular, Englert, Schwinger (who shared the Nobel prize with Feynman and Tomonaga), Scully, and Walther showed that excited atoms passing through a microwave cavity can leave a photon in the cavity without suffering overall recoil (*10–12*).

Thus, by using the wave-like properties of the atom and placing a cavity in front of each slit, we could obtain which-way information (photon left in one cavity or the other). Furthermore, it is easy to envision ways to erase this information and regain fringes.

Again, spirited debate and decisive experiments followed. In this regard, the beautiful experiments of Dürr, Nonn, and Rempe are summarized in the following quotations taken from (*13*), where they note that the party line has it that (p. 33)

> *[If] a which-way detector is employed to determine the particle's path, the interference pattern is destroyed. This is usually explained in terms of Heisenberg's uncertainty principle.*

They further note (p. 33)

> *However, Scully et al. (10, 11) have recently proposed a new gedanken experiment, where the loss of the interference pattern in an atomic beam is not related to Heisenberg's position-momentum uncertainty relation.*

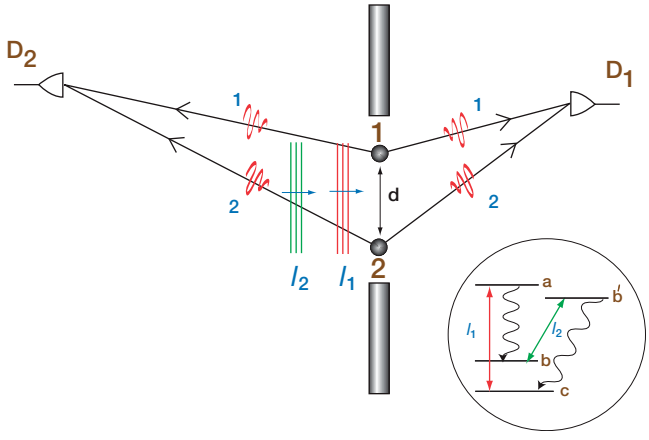


Fig. 5. Two atoms of the type shown located at sites 1 and 2 are separated by a distance *d*. Incident pulse sequence l_1 and l_2 leads to emission of γ and ϕ photons as in quantum eraser. The γ and ϕ photons are detected at D_1 and D_2, respectively. An intensity-intensity correlation yields resolution beyond the classical limit.

This stirred up considerable controversy; to wit (p. 33):

> *Nevertheless, the gedanken experiment of Scully et al. (10, 11) was criticized by Storey et al. (14), who argued that the uncertainty relation always enforces recoil kicks sufficient to wash out the fringes. This started a controversial discussion about the following question: 'Is complementarity more fundamental than the uncertainty principle?'*

They summarize their results and conclusions as follows (p. 33):

> *Here we report a which-way experiment in an atom interferometer in which the 'back action' of path detection on the atom's momentum is too small to explain the disappearance of the interference pattern. We attribute it instead to correlations between which-way detector and atomic motion, rather than to the uncertainty principle.*

Entanglement Quantum Erasers

The preceding discussions showed how quantum eraser can be used to retrieve interference by means of tag ancilla photons $|\phi_\pm\rangle$ going with $|\gamma_\pm\rangle$ fringe and antifringe states. Garisto and Hardy (*15*) invented an interesting new class of quantum erasers, called the disentanglement eraser. These consist of at least three subsystems *A*, *B*, and *T*. The *AB* subsystem is prepared in entangled states of the type

$$|\psi_\pm\rangle = \frac{1}{\sqrt{2}} \times (|0_A, 1_B\rangle \pm |1_A, 0_B\rangle)$$

They then showed how tag states $|\phi_\pm\rangle$ can be used to remove or restore the entanglement. Thus, an outcome $|\phi_+\rangle$ for the tagged state restores the original state $|\psi_+\rangle$ for the *AB* subsystem, whereas the outcome $|\phi_-\rangle$ yields $|\psi_-\rangle$. Thus, a measurement of the tagging qubit restores the entangled state.

An implementation of such an eraser has been demonstrated in nuclear magnetic resonance systems (*16*). Furthermore, a cavity quantum electrodynamics–based implementation has been proposed in (*17*), which provides new insights into quantum teleportation and/or quantum dense coding.

Quantum Kaon Erasers

In a recent article (*18*), Bramon, Garbarino and Hiesmayr have extended these ideas to nuclear physics and showed that an entangled pair of neutral kaons can also display quantum erasure. In their set-up, strangeness oscillations between K^0 and $\overline{K}^0$ display oscillatory (wave-like) behavior and the alternative (which-path like) representation involving eigenstates of mass. The latter representations are called K_S and K_L because they live for about 10^{-10} and 10^{-8} s in free space. As indicated in Fig. 4, the oscillator involves a π incident on plate 1 produces a K^0 that has oscillations when expressed in terms of the K_S and K_L representation. Upon passing through the second plate, only K^0 emerges and this shows typical interference phenomena as indicated. Thus, the kaon oscillations are produced by changing the distance between the two plates. To summarize, then, with no plates we have which-way information associated with decay into two or three π particles. With the plates in place, nucleonic interactions occur, and we can observe oscillatory fringe information. Quantum eraser is achieved by using the entangled state produced by $p\,\overline{p}$ collisions.

Quantum Imaging via Quantum Eraser

Quantum interferometry using entangled photons, as in the paradigm of the quantum eraser, can be used to exceed the resolution limit of classical wave optics. The key resource needed is the ability to jointly measure and correlate the detection of two photons, as described by the intensity correlation function $G^{(2)}$. In the second order interferometry based on photon pairs, the resolution in the measurement of the distance d between the photon sources (Fig. 5) can be potentially improved by as much as an order of magnitude. In order to understand this enhanced resolution, we consider the Scully-Drühl quantum eraser configuration of Fig. 5. The atom of the type shown in Fig. 2C is first excited by a pulse l_1 of center frequency ν_p and much later by a pulse l_2 at frequency ν_d. A γ photon as well as a ϕ photon are emitted either by atom 1 or atom 2 that are detected by detectors D_1 and D_2. The photon-photon correlation function factorizes. The interference pattern observed by moving detector D_1 (and requiring a correlation with detector D_2) is now governed by $k_\gamma + k_\phi \approx 2k$, i.e., the effective radiation wavelength is now $\lambda/2$, leading to an immediate two-fold enhancement beyond the classical limit. In fact, Scully has shown that further improvement results from a more detailed analysis, leading to the possibility of an order of magnitude improvement of resolution (*19*).

References and Notes

1. M. O. Scully, K. Drühl, *Phys. Rev. A.* **25**, 2208 (1982).
2. S. P. Walborn, M. O. T. Cunha, S. Pádua, C. H. Monken, *Am. Sci.* **91**, 336 (2003).
3. S. Begley, *Newsweek*, 19 June 1995, p. 67.
4. B. Greene, *The Fabric of the Cosmos* (Alfred A. Knopf, New York, 2004).
5. R. P. Feynman, R. Leighton, M. Sands, *The Feynman Lectures on Physics, Vol. III* (Addison Wesley, Reading, MA, 1965).
6. Y.-H. Kim, R. Yu, S. P. Kulik, Y. Shih, M. O. Scully, *Phys. Rev. Lett.* **84**, 1 (2000).
7. U. Mohrhoff, *Am. J. Phys.* **64**, 1468 (1996).
8. U. Mohrhoff, *Am. J. Phys.* **67**, 330 (1999).
9. Mathematically we can understand the essential results of the Scully-Drühl quantum eraser by first realizing that the photon state emitted by the atoms located at sites 1 and 2 is given by

$$|\psi_0\rangle = \frac{1}{\sqrt{2}}(|\gamma_1\rangle|\phi_1\rangle + |\gamma_2\rangle|\phi_2\rangle)$$

i.e., either the photon pair γ_1, ϕ_1 is emitted by the atom located at site 1 or the pair γ_2, ϕ_2 is emitted by the atom at site 2. Thus if the ϕ photon is detected by D_3, the quantum state reduces to $|\gamma_1\rangle$. A similar result is obtained for the ϕ photon detection by the detector D_4. This is the situation when the which-path information is available and the sorted data yields no interference fringes. The physics behind the retrieval of the fringes is made clear by rewriting the state $|\psi_0\rangle$ as

$$|\psi_0\rangle = \frac{1}{\sqrt{2}}(|\gamma_+\rangle|\phi_+\rangle + |\gamma_-\rangle|\phi_-\rangle)$$

where $\gamma_\pm$ and $\phi_\pm$ are the symmetric and antisymmetric combinations.

$$|\gamma_\pm\rangle = \frac{1}{\sqrt{2}}(|\gamma_1\rangle + |\gamma_2\rangle)$$

$$|\phi_\pm\rangle = \frac{1}{\sqrt{2}}(|\phi_1\rangle + |\phi_2\rangle)$$

The state of the ϕ photon after passage through the beam splitter B is either $|\phi_+\rangle$ or $|\phi_-\rangle$. Thus, a click at detectors D_1 or D_2, reduces the state of the γ photon to $|\gamma_+\rangle$ or $|\gamma_-\rangle$, respectively, leading to a retrieval of the interference fringes.

10. M. O. Scully, B.-G. Englert, H. Walther, *Nature* **351**, 111 (1991).
11. B.-G. Englert, J. Schwinger, M. O. Scully, *Found. Phys.* **18**, 1045 (1988).
12. M. O. Scully, M. S. Zubairy, *Quantum Optics* (Cambridge, London, 1997).
13. S. Durr, T. Nonn, G. Rempe, *Nature* **395**, 33 (1998).
14. P. Storey, S. Tan, M. Collett, D. Walls, *Nature* **367**, 626 (1994).
15. R. Garisto, L. Hardy, *Phys. Rev. A.* **60**, 827 (1999).
16. G. Teklemariam, E. M. Fortunato, M. A. Pravia, T. F. Havel, D. G. Cory, *Phys. Rev. Lett.* **86**, 5845 (2001).
17. M. S. Zubairy, G. S. Agarwal, M. O. Scully, *Phys. Rev. A.* **70**, 012316 (2004).
18. A. Bramon, G. Garbarino, B. C. Hiesmayr, *Phys. Rev. Lett.* **92**, 020405 (2004).
19. M. O. Scully, unpublished results.

20. We thank E. Fry, A. Muthukrishnan, R. Ooi, and A. Patnaik for their help in the preparation of this manuscript. We also gratefully acknowledge support from U.S. Air Force Office of Scientific Research, Defense Advanced Research Projects Agency, and Texas A&M University's Telecommunication and Informatics Task Force initiative.

10.1126/science.1107787

REVIEW

Astrophysical Observations: Lensing and Eclipsing Einstein's Theories

Charles L. Bennett

Albert Einstein postulated the equivalence of energy and mass, developed the theory of special relativity, explained the photoelectric effect, and described Brownian motion in five papers, all published in 1905, 100 years ago. With these papers, Einstein provided the framework for understanding modern astrophysical phenomena. Conversely, astrophysical observations provide one of the most effective means for testing Einstein's theories. Here, I review astrophysical advances precipitated by Einstein's insights, including gravitational redshifts, gravitational lensing, gravitational waves, the Lense-Thirring effect, and modern cosmology. A complete understanding of cosmology, from the earliest moments to the ultimate fate of the universe, will require developments in physics beyond Einstein, to a unified theory of gravity and quantum physics.

Einstein's 1905 theories form the basis for much of modern physics and astrophysics. In 1905, Einstein postulated the equivalence of mass and energy (*1*), which led Sir Arthur Eddington to propose (*2*) that stars shine by converting their mass to energy via $E = mc^2$, and later led to a detailed understanding of how stars convert mass to energy by nuclear burning (*3*, *4*). Einstein explained the photoelectric effect by showing that light quanta are packets of energy (*5*), and he received the 1921 Nobel Prize in physics for this work. With the photoelectric effect, astronomers determined that ultraviolet photons emitted by stars impinge on interstellar dust and overcome the work function of the grains to cause electrons to be ejected. The photoelectrons emitted by the dust grains excite the interstellar gas, including molecules with molecular sizes of ~1 nm, as estimated by Einstein in 1905 (*6*). Atoms and molecules emit spectral lines according to Einstein's quantum theory of radiation (*7*). The concepts of spontaneous and stimulated emission explain astrophysical masers and the 21-cm hydrogen line, which is observed in emission and absorption. The interstellar gas, which is heated by starlight, undergoes Brownian motion, as also derived by Einstein in 1905 (*8*).

Two of Einstein's five 1905 papers introduced relativity (*1*, *9*). By 1916, Einstein had generalized relativity from systems moving with a constant velocity (special relativity) to accelerating systems (general relativity).

Space beyond Earth provides a unique physics laboratory of extreme pressures and temperatures, high and low energies, weak and strong magnetic fields, and immense dimensions that cannot be reproduced in laboratories or under terrestrial conditions. The extreme astrophysical environments

Department of Physics and Astronomy, The Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA. E-mail: cbennett@jhu.edu

Quantum Imaging via Quantum Eraser

Quantum interferometry using entangled photons, as in the paradigm of the quantum eraser, can be used to exceed the resolution limit of classical wave optics. The key resource needed is the ability to jointly measure and correlate the detection of two photons, as described by the intensity correlation function $G^{(2)}$. In the second order interferometry based on photon pairs, the resolution in the measurement of the distance d between the photon sources (Fig. 5) can be potentially improved by as much as an order of magnitude. In order to understand this enhanced resolution, we consider the Scully-Drühl quantum eraser configuration of Fig. 5. The atom of the type shown in Fig. 2C is first excited by a pulse l_1 of center frequency ν_p and much later by a pulse l_2 at frequency ν_d. A γ photon as well as a ϕ photon are emitted either by atom 1 or atom 2 that are detected by detectors D_1 and D_2. The photon-photon correlation function factorizes. The interference pattern observed by moving detector D_1 (and requiring a correlation with detector D_2) is now governed by $k_\gamma + k_\phi \approx 2k$, i.e., the effective radiation wavelength is now $\lambda/2$, leading to an immediate two-fold enhancement beyond the classical limit. In fact, Scully has shown that further improvement results from a more detailed analysis, leading to the possibility of an order of magnitude improvement of resolution (*19*).

References and Notes

1. M. O. Scully, K. Drühl, *Phys. Rev. A.* **25**, 2208 (1982).
2. S. P. Walborn, M. O. T. Cunha, S. Pádua, C. H. Monken, *Am. Sci.* **91**, 336 (2003).
3. S. Begley, *Newsweek*, 19 June 1995, p. 67.
4. B. Greene, *The Fabric of the Cosmos* (Alfred A. Knopf, New York, 2004).
5. R. P. Feynman, R. Leighton, M. Sands, *The Feynman Lectures on Physics, Vol. III* (Addison Wesley, Reading, MA, 1965).
6. Y.-H. Kim, R. Yu, S. P. Kulik, Y. Shih, M. O. Scully, *Phys. Rev. Lett.* **84**, 1 (2000).
7. U. Mohrhoff, *Am. J. Phys.* **64**, 1468 (1996).
8. U. Mohrhoff, *Am. J. Phys.* **67**, 330 (1999).
9. Mathematically we can understand the essential results of the Scully-Drühl quantum eraser by first realizing that the photon state emitted by the atoms located at sites 1 and 2 is given by

$$|\psi_0\rangle = \frac{1}{\sqrt{2}}(|\gamma_1\rangle|\phi_1\rangle + |\gamma_2\rangle|\phi_2\rangle)$$

i.e., either the photon pair γ_1, ϕ_1 is emitted by the atom located at site 1 or the pair γ_2, ϕ_2 is emitted by the atom at site 2. Thus if the ϕ photon is detected by D_3, the quantum state reduces to $|\gamma_1\rangle$. A similar result is obtained for the ϕ photon detection by the detector D_4. This is the situation when the which-path information is available and the sorted data yields no interference fringes. The physics behind the retrieval of the fringes is made clear by rewriting the state $|\psi_0\rangle$ as

$$|\psi_0\rangle = \frac{1}{\sqrt{2}}(|\gamma_+\rangle|\phi_+\rangle + |\gamma_-\rangle|\phi_-\rangle)$$

where $\gamma_\pm$ and $\phi_\pm$ are the symmetric and antisymmetric combinations.

$$|\gamma_\pm\rangle = \frac{1}{\sqrt{2}}(|\gamma_1\rangle + |\gamma_2\rangle)$$

$$|\phi_\pm\rangle = \frac{1}{\sqrt{2}}(|\phi_1\rangle + |\phi_2\rangle)$$

The state of the ϕ photon after passage through the beam splitter B is either $|\phi_+\rangle$ or $|\phi_-\rangle$. Thus, a click at detectors D_1 or D_2, reduces the state of the γ photon to $|\gamma_+\rangle$ or $|\gamma_-\rangle$, respectively, leading to a retrieval of the interference fringes.

10. M. O. Scully, B.-G. Englert, H. Walther, *Nature* **351**, 111 (1991).
11. B.-G. Englert, J. Schwinger, M. O. Scully, *Found. Phys.* **18**, 1045 (1988).
12. M. O. Scully, M. S. Zubairy, *Quantum Optics* (Cambridge, London, 1997).
13. S. Durr, T. Nonn, G. Rempe, *Nature* **395**, 33 (1998).
14. P. Storey, S. Tan, M. Collett, D. Walls, *Nature* **367**, 626 (1994).
15. R. Garisto, L. Hardy, *Phys. Rev. A.* **60**, 827 (1999).
16. G. Teklemariam, E. M. Fortunato, M. A. Pravia, T. F. Havel, D. G. Cory, *Phys. Rev. Lett.* **86**, 5845 (2001).
17. M. S. Zubairy, G. S. Agarwal, M. O. Scully, *Phys. Rev. A.* **70**, 012316 (2004).
18. A. Bramon, G. Garbarino, B. C. Hiesmayr, *Phys. Rev. Lett.* **92**, 020405 (2004).
19. M. O. Scully, unpublished results.

20. We thank E. Fry, A. Muthukrishnan, R. Ooi, and A. Patnaik for their help in the preparation of this manuscript. We also gratefully acknowledge support from U.S. Air Force Office of Scientific Research, Defense Advanced Research Projects Agency, and Texas A&M University's Telecommunication and Informatics Task Force initiative.

10.1126/science.1107787

REVIEW

Astrophysical Observations: Lensing and Eclipsing Einstein's Theories

Charles L. Bennett

Albert Einstein postulated the equivalence of energy and mass, developed the theory of special relativity, explained the photoelectric effect, and described Brownian motion in five papers, all published in 1905, 100 years ago. With these papers, Einstein provided the framework for understanding modern astrophysical phenomena. Conversely, astrophysical observations provide one of the most effective means for testing Einstein's theories. Here, I review astrophysical advances precipitated by Einstein's insights, including gravitational redshifts, gravitational lensing, gravitational waves, the Lense-Thirring effect, and modern cosmology. A complete understanding of cosmology, from the earliest moments to the ultimate fate of the universe, will require developments in physics beyond Einstein, to a unified theory of gravity and quantum physics.

Einstein's 1905 theories form the basis for much of modern physics and astrophysics. In 1905, Einstein postulated the equivalence of mass and energy (*1*), which led Sir Arthur Eddington to propose (*2*) that stars shine by converting their mass to energy via $E = mc^2$, and later led to a detailed understanding of how stars convert mass to energy by nuclear burning (*3*, *4*). Einstein explained the photoelectric effect by showing that light quanta are packets of energy (*5*), and he received the 1921 Nobel Prize in physics for this work. With the photoelectric effect, astronomers determined that ultraviolet photons emitted by stars impinge on interstellar dust and overcome the work function of the grains to cause electrons to be ejected. The photoelectrons emitted by the dust grains excite the interstellar gas, including molecules with molecular sizes of ~1 nm, as estimated by Einstein in 1905 (*6*). Atoms and molecules emit spectral lines according to Einstein's quantum theory of radiation (*7*). The concepts of spontaneous and stimulated emission explain astrophysical masers and the 21-cm hydrogen line, which is observed in emission and absorption. The interstellar gas, which is heated by starlight, undergoes Brownian motion, as also derived by Einstein in 1905 (*8*).

Two of Einstein's five 1905 papers introduced relativity (*1*, *9*). By 1916, Einstein had generalized relativity from systems moving with a constant velocity (special relativity) to accelerating systems (general relativity).

Space beyond Earth provides a unique physics laboratory of extreme pressures and temperatures, high and low energies, weak and strong magnetic fields, and immense dimensions that cannot be reproduced in laboratories or under terrestrial conditions. The extreme astrophysical environments

Department of Physics and Astronomy, The Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA. E-mail: cbennett@jhu.edu

enable us to test Einstein's theories, and the general theory of relativity (*10*) has been tested and applied almost exclusively in the astrophysical arena.

Gravity as Geometry

Newton formulated the first universal laws of gravity and motion. Using Newton's laws, astronomers can predict how moons will orbit planets. Newton, however, could not understand how a moon would sense the presence of a planet and wrote, ''I have not been able to discover the cause of those properties of gravity from phenomena, and I frame no hypotheses…'' (*11*). Einstein's theory of general relativity answered Newton's question: Mass causes space-time curvature. A planet bends the space around it, and its moon moves under the influence of that bent space.

Einstein encouraged astronomers to measure the positions of stars near the Sun to see how much the light was deflected. During an eclipse, it is possible to see stars near the occulted solar disk. An accurate measurement of the apparent position of one of these stars would provide a test of Einstein's idea (Fig. 1). Eddington traveled to Principe Island in western Africa for the 29 May 1919 eclipse, and he claimed to have seen evidence of the starlight shift predicted by Einstein. Today, we no longer are constrained to viewing optical light during eclipses; high-resolution radio interferometers, for example, can measure the position of quasar light that passes near the Sun. These recent measurements (*12*) provide more compelling support for relativity. Irwin Shapiro recognized that when light traverses curved space, such as near the Sun, it will arrive at its destination at a later time than light that passes through unperturbed flat space (*13*). Shapiro used radar ranging to planets and spacecraft to determine that the Sun's bending of space adds about 200 μs to the round-trip travel time of a radio signal passing near the edge of the Sun. The Shapiro time delay was recently confirmed by tracking the Cassini spacecraft during its flight to Saturn (*14*). The measurement agreed with relativity to 1 part in 10^5.

Of the planets in the solar system, Mercury's orbit is most affected by the Sun's gravity (Fig. 1). The radius of Mercury's elliptical orbit (with an eccentricity of 0.206) varies from 46 to 70 million km. The measurement of the advance of the perihelion (the closest approach of an orbit to the Sun) of Mercury from the Earth is 5599 arc-sec per century, of which 5025 arc-sec per century are caused by the precession of Earth's equinoxes. The remaining 574 arc-sec per century do not match the Newtonian prediction that the axial direction of Mercury's elliptical orbit about the Sun should precess, advancing by 531 arc-sec per century, caused by perturbations from the other planets. In 1855, Le Verrier hypothesized the existence of an undetected planet, Vulcan, between Mercury and the Sun, to explain this discrepancy. Instead, general relativity accounts for the remaining 43 arc-sec per century and for the more constraining recent data acquired from radio tracking of the Mars landers.

During the 1930s, Fritz Zwicky determined that clusters of galaxies have 10 times more mass than their light suggests, and termed the unseen mass ''dark matter.'' He suggested that the matter would cause space to curve and act as a gravitational lens, and that the clusters could be ''weighed'' by the lensing effects (*15*, *16*). The suggestion was not to test Einstein's definition of gravity, but to use it as a tool. Gravitationally lensed systems confirm Zwicky's assertion that clusters are dominated by dark matter (Fig. 2). Observations of the rotation curves of individual spiral galaxies also require large amounts of dark matter in halos around the galaxies (*17*).

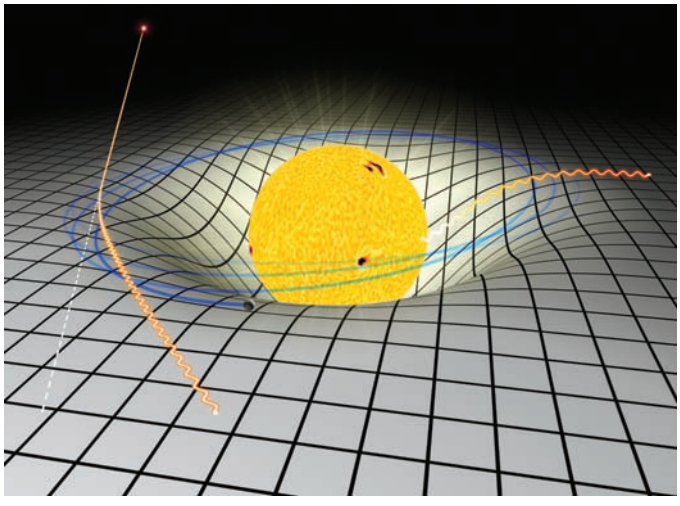


Fig. 1. A mass like our Sun bends space. On the right, light loses energy as it climbs the gravitational potential, undergoing a gravitational redshift. On the left, the path of distant starlight is deflected (or gravitationally lensed) by the curved space. Because of the relativistic effects of the curved space, the deflection angle is twice that predicted by Newtonian calculation. The advance of the perihelion of the orbit of the planet Mercury (seen in blue) is slightly altered by relativistic effects, as discussed in the text. [Figure by Theophilus Britt Griswold]

Gravitational Redshifts

A central tenet of relativity is Einstein's equivalence principle: It is impossible to distinguish between a local gravitational field and an equivalent uniform acceleration in a sufficiently small region of space-time. The gravitational redshift follows directly from this equivalence principle. Light that propagates through a gravitational field changes wavelength according to $\lambda/\lambda_0 = \Delta\Phi/c^2$, where λ is the shifted wavelength, λ_0 is the rest wavelength, c is the speed of light, and $\Delta\Phi$ is the change in the Newtonian gravitational potential (valid in the weak gravitational field limit). In strong gravitational fields that are a radial distance r from a mass M (where r is defined as a circumference divided by 2π), $\lambda/\lambda_0 = (1 - 2GM/c^2r)^{1/2}$, where G is the gravitational constant. The observed wavelength goes to zero at the Schwarzschild radius, $R_S = 2GM/c^2$ (the event horizon). The gravitational redshift was confirmed to 10% accuracy in 1960 by using the transmission of a very narrow 14.4-keV transition from the isotope ^{57}Fe across the 22.6-m height of the Jefferson Physical Laboratory building at Harvard (*18*). Later refinements of these Pound-Rebka-Snider experiments reached 1% accuracy. Vessot *et al*. (*19*) used a hydrogen maser clock on a rocket launched to an altitude of 10^4 km to measure the gravitational redshift to an accuracy of 7×10^{-5}.

Gravitational Waves

Maxwell's equations explain how an accelerating charged particle produces electromagnetic radiation. Analogously, general relativity predicts that an accelerating mass produces gravitational radiation. This radiation has not yet been measured directly, but it has been detected indirectly. From 1974 to 1983, Hulse and Taylor observed the pulsar (*20*) PSR1913+16, which orbits a 1.4–solar mass ($M_\odot$) star with a period of only 7.75 hours (*21*).

According to general relativity, these two massive bodies in rapid motion produce gravitational waves, and the emission of these waves should cause the binary system to lose energy. Indeed, the period of the pulsar's orbit was observed to be increasing at a rate of 76 millionths of a second per year, consistent (within 0.02%) with the expected energy loss caused by gravitational waves (*22*).

Considerable effort is now underway to make direct measurements of 10- to 1000-Hz gravitational radiation with the Laser Interferometer Gravitational Wave Observatory (LIGO). The Laser Interferometer Space Antenna (LISA), a planned three-spacecraft

constellation, will measure gravitational waves in the range from 0.0001 to 1 Hz. LIGO and LISA (Fig. 3) will seek to detect gravitational waves from orbits, coalescences, and collisions of strong-field objects such as neutron stars and black holes. Their objective is to probe gravity in systems where the ratio of the gravitational binding energy to rest mass is in the range of 0.1 to 1, which is many orders of magnitude beyond any current measurement.

The Lense-Thirring Effect

General relativity predicts the existence of black holes and compelling evidence has been found for their existence in many galaxies (*23*). Stars that exhaust their hydrogen supply cannot start burning carbon unless they have more than 4 $M_\odot$. Stars with less mass will collapse until the contraction is halted by electron degeneracy, according to the Pauli exclusion principle. These contracted stars are called white dwarfs. If the mass of the collapsing star is greater than 1.44 $M_\odot$, then electrons and protons will combine to form neutrons, and the star will collapse further until it is supported by the neutron degeneracy of the Pauli exclusion principle. These more massive contracted stars are called neutron stars. If the mass is greater than 2 to 3 $M_\odot$, even the neutron degeneracy cannot support the mass. With nothing to stop it, the star will collapse to become a black hole. The strong gravitational fields associated with neutron stars and black holes provide a laboratory for testing general relativity.

For a gyroscope near a rotating black hole, distant stars provide an inertial reference frame, but the black hole defines a different reference frame. The closer the gyroscope is to the black hole, the more it will be affected by the black hole's reference frame. The gyroscope will precess while trying to follow the rotation of the black hole, but it will lag behind because of its fidelity to the distant stars. In 1918, Joseph Lense and Hans Thirring introduced this problem, which does not arise in Newtonian gravity. The Lense-Thirring precession can be thought of as the dragging of inertial frames around rotating masses. The rotation twists the surrounding space, perturbing the orbits of nearby masses.

Possible x-ray detections of frame dragging around neutron stars and black holes have been suggested (*24*). Unfortunately, attempts to measure the Lense-Thirring effect around black holes are often confounded by the complexities of the dynamics of the hot gas in their accretion disks. Earth's gravitational field, although much weaker, should also cause frame dragging. A detection of the Lense-Thirring effect has been claimed with 10% accuracy based on the precession of the orbital planes of the two LAser GEOdynamics Satellites (LAGEOS) (*25*). LAGEOS-1 and LAGEOS-2 are 0.6-m-diameter aluminum spherical shells in Earth orbit, each covered with 426 passive retroreflectors.

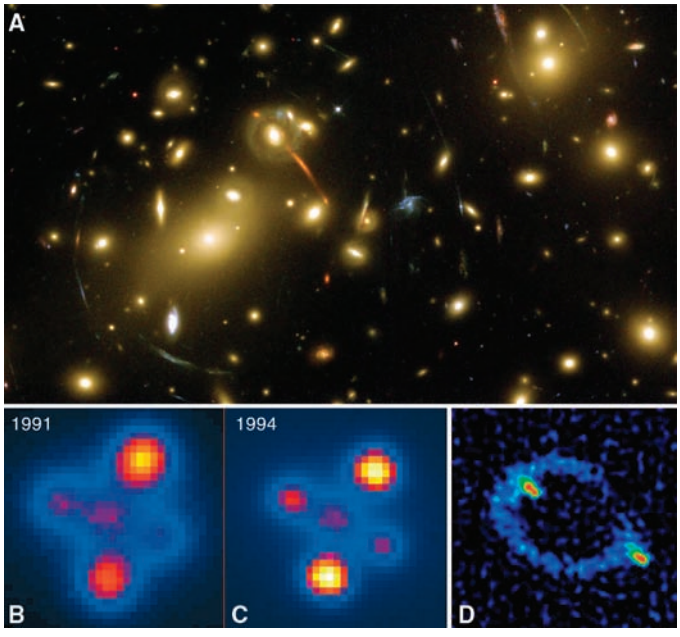


Fig. 2. The distribution of dark matter in space can be deduced from gravitational lensing effects. (**A**) An HST picture of the galaxy cluster Abell 2218 shows distorted images of the background light, including arcs, contorted galaxy shapes, and multiple images of single objects. [Courtesy of NASA by A. Fruchter and the HST Early Release Observations Team] With sources of variable luminosity, differential propagation time delays become apparent in the multiple images, such as with the 3-year change in the Einstein cross (**B** and **C**). In a coaligned system (source to lens, mass to observer), the source appears as an Einstein ring (**D**). [Courtesy of G. Lewis (Institute of Astronomy) and M. Irwin (Royal Greenwich Observatory) from the William Hershel Telescope and J. N. Hewitt and E. L. Turner from the National Radio Astronomy Observatory/Associated Universities, Inc.]

Analogous to an accelerating charge producing a magnetic field, an accelerating mass produces certain ''gravitomagnetic'' effects. Three gravitomagnetic effects are: (i) the precession of a gyroscope in orbit about a rotating mass, (ii) the precession of orbital planes in which a mass orbiting a large rotating body constitutes a gyroscopic system whose orbital axis will precess, and (iii) the precession of the pericenter of the orbit of a test mass about a massive rotating object. NASA's Gravity Probe B (GPB) seeks to detect and measure the spin-spin interaction between its own orbiting gyroscope and the spinning Earth (frame dragging) and the combination of the effects of spin-orbit coupling and space-time curvature known as geodetic (or Thomas) precession.

Big Bang Nucleosynthesis

Big Bang nucleosynthesis calculations depend on the relation between temperature and the rate of expansion of the universe. This relation is specified by the Friedmann equations, which in turn are derived from Einstein's field equations. Thus, the degree to which Big Bang nucleosynthesis matches the measured abundance ratios of the light chemical elements is also a test of Einstein's equations. In 1946, George Gamow wrote that it was ''generally agreed … that the relative abundance of various chemical elements were determined by physical conditions existing in the universe during the early stages of its expansion, when the temperature and density were sufficiently high to secure appreciable reaction-rates for the light as well as for the heavy elements.'' (*26*). Gamow gave the problem to his graduate student, Ralph Alpher, to work out. Alpher's Ph.D. thesis and a resulting journal paper (*27*) appeared in 1948, showing in detail that the lightest elements (with atomic numbers ≤ 4) would be formed in the early universe. Gamow was wrong about the heavier elements, which we now know were created later in stellar processes (*28*).

Our Dynamic Universe

General relativity is at the heart of modern cosmological models. Einstein's field equation is $R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi G T_{\mu\nu}$, where $R_{\mu\nu}$ is the Ricci tensor, R is the Ricci scalar, G is Newton's gravitational constant, $g_{\mu\nu}$ is the metric tensor, and $T_{\mu\nu}$ is the stress-energy tensor. The left-hand side of the equation describes the curvature of space and the right-hand side describes the content of space. Thus, the equation reflects the equivalence of matter and curved space. When Einstein applied his field equation to cosmology, he found a problem. His equations implied an unstable (evolving) universe. To produce a stable universe, he needed to add a parameter to his equation in order to overcome the attractive gravity of all of the mass. In 1917, he introduced what he called the cosmological constant as a repulsive antigravity term into the field equation. The field equation with Einstein's cosmological con-

stant term, Λ, becomes $R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi GT_{\mu\nu}$. The cosmological constant can be thought of as the energy of the vacuum (that is, how much ''nothing'' weighs) and is a constant of nature, with no specified value. Einstein fine-tuned its value to balance exactly the tendency of mass in the universe to collapse gravitationally.

Now Einstein had a static finite universe, but with a second fundamental gravitational constant of nature. Not only was this a fine-tuned solution, but it was unstable to perturbations; the tiniest of fluctuations would cause the universe to shrink or to expand.

This solution was also based on the erroneous assumption that the universe is static. In 1929, Edwin Hubble determined that the recession velocities of galaxies (determined by Doppler shifts) are proportional to their distances. Assuming that we do not occupy a special place in the universe, Hubble concluded that the universe is expanding. Given an expanding universe and not a static one, Einstein concluded that his cosmological constant was unnecessary.

For Hubble, Cepheid variable stars provided the key connection to obtain the distance to galaxies. The recession velocities of galaxies were determined by the Doppler shift of their light. More recently, Type Ia supernovae light curves have been used for distance determinations. In an attempt to measure the slowing of the expansion of the universe by using supernovae, astronomers discovered that the expansion of the universe is accelerating, not decelerating (*29*, *30*). Einstein's cosmological constant has reemerged as one of many candidate explanations for the accelerated expansion.

The Cosmic Microwave Background

In working through Big Bang nucleosynthesis in 1949, Alpher and Herman estimated (*31*) that blackbody radiation with ''a temperature on the order of 5 K'' would remain today, ''interpreted as the background temperature which would result from the universal expansion alone.'' Arno Penzias and Robert Wilson won the 1978 Nobel Prize for their 1965 detection of this afterglow of the Big Bang. The radiation has now been precisely and accurately characterized by using the Cosmic Background Explorer (COBE) satellite as a black body with a radiation temperature of 2.725 ± 0.001 K (Fig. 4A) (*32*). The blackbody spectrum provided evidence that the radiation had been in thermal equilibrium with matter earlier in the history of the universe, as predicted within the Big Bang model.

Penzias and Wilson found the cosmic microwave background (CMB) to be uniform across the sky to within the precision of their measurements. Further measurements were performed in order to detect the nonuniformity that was expected because the gravitational seeds of the clumpy structures we see in the sky today should be associated with small temperature variations across the sky in the CMB. Year after year, experimentalists improved technology to reach ever fainter upper limits on the amplitude of these temperature fluctuations, and the lack of detectable fluctuations began to be a problem. Without large enough primordial gravitational fluctuations, there wouldn't be sufficient time for gravity to cause the growth of the current structure of the universe. A solution to the structure problem came with the introduction of cold dark matter (CDM) (*33*). Although baryons may not always emit much light and may be thought of as dark matter, CDM is fundamentally different because CDM particles do not interact with light.

In the early universe, when it was too hot for electrons to bind to hydrogen atoms, there was a plasma of photons, electrons, baryons, and CDM. The photons would frequently scatter off of the free electrons. The seeds of the structure of the universe were also present in the form of primordial gravitational perturbations. The CDM would congregate in regions of gravitational potential minima. The combined gravitational pull of the primordial fluctuation and the CDM would attract the baryonic matter, but unlike the CDM, the baryonic matter would experience enormous radiation pressure that would prevent the baryons from clumping. The opposing forces of gravitational attraction and radiation pressure induced adiabatic oscillations in the photon-baryon fluid.

This occurred within the general relativistic expansion and cooling of the universe. As the universe expanded, the radiation pressure dropped and the universe became matter-dominated. At that point, the baryons also participated in the growth of structures. When the universe cooled sufficiently for electrons to bind to protons to form neutral hydrogen (~3000 K), the photons no longer scattered. These CMB photons we now observe carry with them the signature of the epoch of last scattering, called decoupling. On an angular scale greater than 2°, the photons have been gravitationally redshifted according to the amplitude of the gravitational potential at decoupling.

In 1992, the 27-year search for the anisotropy in the CMB ended with the COBE discovery of the faint fluctuations in a map of the full sky (Fig. 4B). This discovery supported the basic cosmological picture that structure developed gravitationally in the universe, with the mass dominated by CDM.

The detection of the primordial gravitational anisotropy caused experimentalists to redirect their efforts to the characterization of the statistics of the anisotropy, especially at smaller angular scales. A precise and accurate measurement of the statistics of the anisotropy pattern was predicted to reveal the signature of the adiabatic oscillations, the nature of which depends on the values of key cosmological parameters. CMB balloon-borne (*34*) and ground-based (*35*) experiments collected a wealth of data (*36*) on the nature of the adiabatic oscillations, and the space-borne Wilkinson Microwave Anisotropy Probe (WMAP) mission produced full-sky multi-frequency maps of the CMB (Fig. 4C) that tightly constrain the dynamics, content, and shape of the universe (*37*).

Observational cosmologists have been productive on many other fronts in the past decade. The Hubble Space Telescope (HST) determined the expansion rate of the universe, known as the Hubble constant (*38*). The 2° Field Galaxy Redshift Survey and the Sloan Digital Sky Survey have constrained cosmological parameters (*39*, *40*) by determining the three-dimensional power spectrum based on observations of hundreds of thousands of galaxies.

The combination of the cosmological observations of the past 10 years has led to a highly constrained cosmological model that is consistent with these wide-ranging observations and general relativity: A flat (Euclid-

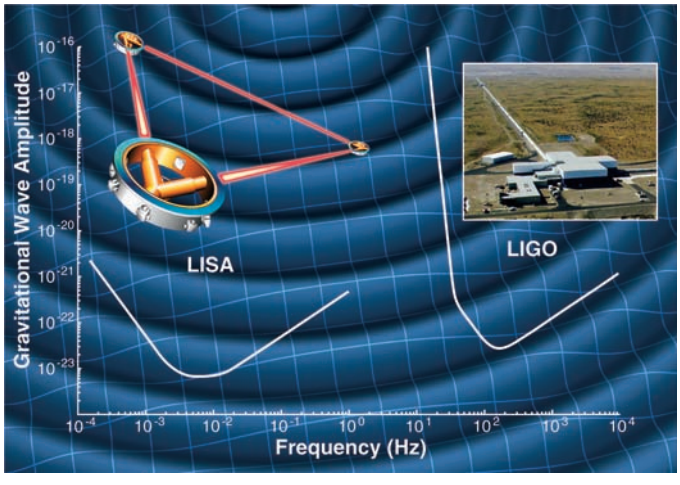


Fig. 3. LIGO is searching for gravitational waves of 3 to 100 ms, which are expected from the collapsing cores of supernovae and from coalescing neutron stars and stellar-mass black holes. LISA is sensitive to gravitational waves with much longer periods and is currently under development, with the goal of detecting events such as the coalescence of supermassive black holes.

ean) universe is dominated by 73% dark energy, 23% CDM, and 4% baryons, and has a Hubble constant of 71 km s^{-1} Mpc^{-1}. Despite this knowledge, we still do not know the specific nature of the dark energy or the CDM, nor do we understand the physics of the origin of the universe.

Inflation

There is evidence that the universe underwent a short early burst of extremely rapid expansion, called inflation. Inflation has been used to explain the so-called flatness, horizon, and monopole problems. We measure space to be nearly flat because the inflationary expansion drove the initial radius of curvature of the universe to become large. We find the CMB temperature to be highly uniform across the sky because, with inflation, regions more than 2° apart on the sky were in causal contact and thus had an opportunity to equilibrate. We observe no magnetic monopoles because the inflationary expansion drove their density to an infinitesimal level.

Inflation is a paradigm, and within the inflationary paradigm there are many specific models. Each model is characterized by a tensor-to-scalar ratio (that is, the ratio of anisotropy power in gravitational waves to adiabatic density perturbations) and a spectral index of spatial fluctuations, both of which are measurable quantities from CMB observations. Inflation generically produces gravitational waves. Inflationary gravitational waves will be exceedingly difficult to detect directly; however, they imprint a signature polarization pattern onto the CMB. The measurement of this pattern could determine the energy level of the inflationary expansion. The energy level of inflation is likely to be at the scale of the grand unified theory (GUT), where the strong and electroweak forces are unified. For GUT-scale inflation, the detection of the polarization patterns should be possible, albeit very difficult. CMB researchers have already begun the search for the primordial gravitational waves. A future NASA mission, the Einstein Inflation Probe, is currently in its early planning stages and is aimed at detecting the primordial gravitational waves by their imprint on the CMB.

Unification

James Clerk Maxwell is famous for presenting the unified theory of electricity and magnetism called electromagnetism. The electromagnetic force was later quantized by Feynman in a theory called quantum electrodynamics (QED). The quantized electromagnetic force was then unified with the weak force by Weinberg, Salam, and Glashow. Analagous to the interactions between photons and charged particles in QED, quarks carry a special kind of "strong charge," arbitrarily called "color," and interact by the strong force in the theory of quantum chromodynamics (QCD). Gross, Wilczek, and Politzer recently shared the 2004 Nobel Prize in physics for their work in QCD. Because the electromagnetic force and the gravitational force both follow $1/r^2$ force laws (the force is inversely proportional to the square of the radial distance r), Einstein felt that these were the most promising forces to try to unify, yet he was unsuccessful. In 1919, Kaluza proposed to unify electromagnetism with gravity by introducing an extra dimension, and in 1926, Klein proposed that the extra dimension be compact. Some string theories predict that spacetime has 10 dimensions. If six dimensions are tightly curled up (compacted), then it would be difficult to establish their existence. We have now eclipsed Einstein with ongoing progress toward a unified quantum field theory and with new ideas for the unification of gravity from superstring and M theory.

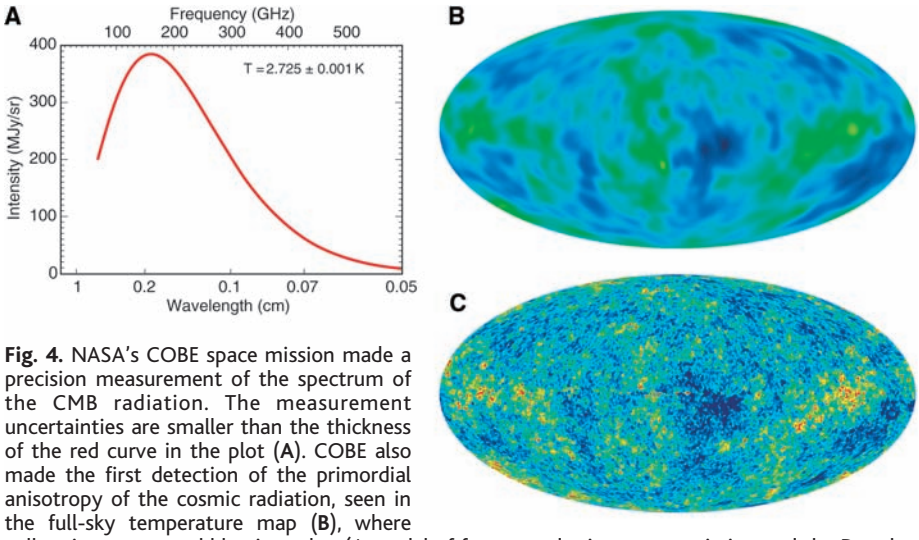


Fig. 4. NASA's COBE space mission made a precision measurement of the spectrum of the CMB radiation. The measurement uncertainties are smaller than the thickness of the red curve in the plot (**A**). COBE also made the first detection of the primordial anisotropy of the cosmic radiation, seen in the full-sky temperature map (**B**), where yellow is warmer and blue is cooler. (A model of foreground microwave emission and the Doppler term have been subtracted.) The primary physical effect is the gravitational redshifting of the light from primordial gravitational potential wells and blueshifting from primordial gravitational potential peaks. (**C**) The higher sensitivity and higher resolution WMAP full-sky map reveals the interactions of the photon-baryon fluid in the early universe, allowing a host of cosmological parameters to be deduced. [Photo courtesy of NASA's COBE and WMAP science teams]

Dark Energy

We now believe that the universe is dominated by a completely unpredicted dark energy. Einstein's cosmological constant could be the dark energy, and all cosmological measurements are consistent with this interpretation. In this scenario, the energy densities of radiation and matter decrease as the universe expands, but the cosmological constant does not. Although the cosmological constant was negligible in the early universe, it now dominates the energy density of the universe (*41*). However, the cosmological constant as a dark energy candidate is unsatisfying because we cannot explain why its observed density is such that it just recently dominates the dynamics of the universe. It is tempting to associate the dark energy with the energy of the vacuum. Physical scaling suggests that the vacuum energy should be enormous, although the cosmological constant is tiny by comparison. Either there is a subtle reason for a near-zero but finite vacuum energy, or the dark energy is not vacuum energy.

The dark energy could be a sign of the breakdown of general relativity. A more general gravity theory may solve the problem, and research continues into various modified gravity theories. Alternatively, the dark energy could be a universally evolving scalar field (*42*).

Many other explanations for the dark energy have also been posited. Although it is often said that we have no idea what the dark energy is, it may be more correct to say that we have too many ideas of what the dark energy might be. And, of course, all of our current ideas may be wrong.

Progress in dark energy research is a high-priority challenge for theoretical and experimental physicists. Currently, two leading approaches for characterizing the dark energy are the use of supernovae light curves as distance indicators and the study of weak gravitational lensing. Weak lensing takes advantage of the contortions of the shapes of distant objects by intervening objects (lenses). Large statistical lensing samples, as a function of redshift, probe both the geometry and dynamics of the universe.

Einstein launched a revolution in physics that continues to this day, and his theories have undergone increasingly rigorous tests. Effects predicted by Einstein's theories have been used as tools to comprehend the universe. In many ways, we are now going beyond Einstein (*43*). On the theoretical frontier, efforts toward the unification of the laws of physics continue. On

the experimental frontier, tireless efforts are underway to detect CDM directly in underground low-background experiments; to produce CDM in particle accelerators; to make measurements to determine the nature of the dark energy (its sound speed, equation of state, and evolution); and to test inflation and determine what happened in the earliest moments of our universe.

References and Notes

1. A. Einstein, *Ann. der Phys.* **18**, 639 (1905).
2. A. S. Eddington, Address to the Mathematical and Physical Science Section (Section A) of the Royal Astronomical Society in Cardiff, and *Nature* **106**, 14 (1920).
3. H. A. Bethe, *Phys. Rev.* **55**, 103 (1939).
4. W. A. Fowler, G. R. Burbidge, E. M. Burbidge, *Astrophys. J.* **122**, 271 (1955).
5. A. Einstein, *Ann. der Phys.* **17**, 132 (1905).
6. A. Einstein, *Ann. der Phys.* **19**, 289 (1906).
7. A. Einstein, *Phys. Z.* **18**, 121 (1917).
8. A. Einstein, *Ann. der Phys.* **17**, 549 (1905).
9. A. Einstein, *Ann. der Phys.* **17**, 891 (1905).
10. A. Einstein, *Ann. der Phys.* **49**, 769 (1916).
11. Letter from Isaac Newton to Robert Hooke, 5 February 1675.
12. E. B. Fomalont, R. A. Sramek, *Phys. Rev. Lett.* **36**, 1475 (1976).
13. I. I. Shapiro, *Phys. Rev. Lett.* **13**, 789 (1964).
14. B. Bertotti, L. less, P. Tortora, *Nature* **425**, 374 (2003).
15. F. Zwicky, *Phys. Rev.* **51**, 290 (1937).
16. F. Zwicky, *Phys. Rev.* **51**, 679 (1937).
17. V. C. Rubin, N. Thonnard, W. K. Ford Jr., *Astrophys. J.* **225**, L107 (1978).
18. R. V. Pound, G. A. Rebka Jr., *Phys. Rev. Lett.* **4**, 337 (1960).
19. R. F. C. Vessot *et al.*, *Phys. Rev. Lett.* **45**, 2081 (1980).
20. A pulsar is a rapidly rotating neutron star that emits highly directional radiation. Pulses are observed as the rotation periodically sweeps the beam across the sky toward the observer, similar to a lighthouse.
21. J. H. Taylor, R. A. Hulse, L. A. Fowler, G. E. Gullahorn, J. M. Rankin, *Astrophys. J.* **206**, L53 (1976).
22. J. H. Taylor, L. A. Fowler, J. M. Weisberg, *Nature* **277**, 437 (1979).
23. M. Begelman, *Science* **300**, 1898 (2003).
24. L. Stella, M. Vietri, *Astrophys. J.* **492**, L59 (1998).
25. I. Ciufolini, E. C. Pavlis, *Nature* **431**, 958 (2004).
26. G. Gamow, *Phys. Rev.* **70**, 572 (1946).
27. R. A. Alpher, H. Bethe, G. Gamow, *Phys. Rev.* **73**, 803 (1948).
28. H. A. Bethe, *Phys. Rev.* **55**, 103 (1939).
29. S. Perlmutter *et al.*, *Astrophys. J.* **517**, 565 (1999).
30. A. G. Riess *et al.*, *Astron. J.* **116**, 1009 (1998).
31. R. A. Alpher, R. C. Herman, *Phys. Rev.* **75**, 1089 (1949).
32. J. C. Mather, D. J. Fixsen, R. A. Shafer, C. Mosier, D. T. Wilkinson, *Astrophys. J.* **512**, 511 (1999).
33. J. R. Bond, G. Efstathiou, *Astrophys. J.* **285**, L45 (1984).
34. For examples of balloon-borne CMB experiments, see the following Web sites: Far Infra-Red Survey (FIRS, http://physics.princeton.edu/~cmb/firs.html); Balloon Observations Of Millimetric Extragalactic Radiation and Geophysics (BOOMERanG, http://cmb.phys.cwru.edu/boomerang/); Millimeter Anisotropy eXperiment Imaging Array (MAXIMA, http://cosmology.berkeley.edu/group/cmb/index.html); QMAP (http://physics.princeton.edu/~cmb/qmap/qmap.html); Medium Scale Anisotropy Measurement and TopHat (MSAM/TopHat, http://topweb.gsfc.nasa.gov/); Archeops (www.archeops.org/index_english.html); Balloon-borne Anisotropy Measurement (BAM, http://cmbr.physics.ubc.ca/bam/experimental.html); ARGO (http://oberon.roma1.infn.it/argo/); Background Emission Anisotropy Scanning Telescope (BEAST/ACE, www.deepspace.ucsb.edu/research/Sphome.htm).
35. For examples of ground-based CMB experiments, see the following Web sites: Arcminute Cosmology Bolometer Array Receiver (ACBAR, http://cosmology.berkeley.edu/group/swlh/acbar/); Advanced Cosmic Microwave Explorer (ACME/HACME, www.deepspace.ucsb.edu/research/Sphome.htm); Antarctic Plateau Anisotropy CHasing Experiment (APACHE, www.bo.iasf.cnr.it/~valenziano/APACHE/apache.htm); Australia Telescope Compact Array (ATCA, www.atnf.csiro.au/research/cmbr/index.html); Cosmic Anisotropy Polarization MAPper (CAPMAP, http://wwwphy.princeton.edu/cosmology/); Cosmic Anisotropy Telescope (CAT, www.mrao.cam.ac.uk/telescopes/cat/index.html); Cosmic Background Imager (CBI, www.astro.caltech.edu/~tjp/CBI/); CosmicMicrowave Polarization at SmallScales (COMPASS, http://cmb.physics.wisc.edu/compass.html); Degree Angular Scale Interferometer (DASI, http://astro.uchicago.edu/dasi/); Princeton I, Q, and U Experiment (PIQUE, http://wwwphy.princeton.edu/cosmology/); Polarization Observations of Large Angular Regions (POLAR, http://cmb.physics.wisc.edu/polar/); Saskatoon/Toco/Mobile Anisotropy Telescope (SK/TOCO/MAT, http://wwwphy.princeton.edu/cosmology/); Python (http://astro.uchicago.edu/cara/research/cmbr/python.html); Tenerife (www.iac.es/project/cmb/rad/index.html); Very Small Array (VSA, www.mrao.cam.ac.uk/telescopes/vsa/index.html); Jodrell Bank; Viper; Owens Valley Radio Observatory (OVRO, www.astro.caltech.edu/~tjp/OVRO-CMB/intrinsic.html).
36. The Legacy Archive for Microwave Background Data Analysis (LAMBDA) data center, available at http://lambda.gsfc.nasa.gov, has links to data and further information about the CMB experiments.
37. C. L. Bennett *et al.*, *Astrophys. J.* **148**, 1 (2003).
38. W. L. Freedman *et al.*, *Astrophys. J.* **553**, 47 (2001).
39. W. J. Percival *et al.*, *Mon. Not. R. Astron. Soc.* **337**, 1068 (2002).
40. M. Tegmark *et al.*, *Phys. Rev.* **D69**, 103501 (2004).
41. P. J. E. Peebles, B. Ratra, *Rev. Mod. Phys.* **75**, 559 (2003).
42. P. Steinhardt, *Philos. Trans. Math. Phys. Eng. Sci.* **361**, 2497 (2003).
43. *Connecting Quarks with the Cosmos: Eleven Science Questions for the New Century* (National Academies Press, Washington, DC, 2003).

44. I thank R. Weiss, E. L. Wright, and G. Hinshaw for helpful comments that led to an improved paper. I am also grateful to B. Griswold for expert graphics assistance. Support for this work was provided by NASA.

10.1126/science.1106444

REVIEW

Inflationary Cosmology: Exploring the Universe from the Smallest to the Largest Scales

Alan H. Guth* and David I. Kaiser*

Understanding the behavior of the universe at large depends critically on insights about the smallest units of matter and their fundamental interactions. Inflationary cosmology is a highly successful framework for exploring these interconnections between particle physics and gravitation. Inflation makes several predictions about the present state of the universe—such as its overall shape, large-scale smoothness, and smaller scale structure—which are being tested to unprecedented accuracy by a new generation of astronomical measurements. The agreement between these predictions and the latest observations is extremely promising. Meanwhile, physicists are busy trying to understand inflation's ultimate implications for the nature of matter, energy, and spacetime.

The scientific community is celebrating the International Year of Physics in 2005, honoring the centennial of Albert Einstein's most important year of scientific innovation. In the span of just a few months during 1905 Einstein introduced key notions that would dramatically change our understanding of matter and energy as well as the nature of space and time. The centennial of these seminal developments offers an enticing opportunity to take stock of how scientists think about these issues today. We focus in particular on recent developments in the field of inflationary cosmology, which draws on a blend of concepts from particle physics and gravitation. The last few years have been a remarkably exciting time for cosmology, with new observations of unprecedented accuracy yielding many surprises. Einstein's legacy is flourishing in the early 21st century.

Inflation was invented a quarter of a century ago, and has become a central ingredient of current cosmological research. Describing dramatic events in the earliest history of our universe, inflationary models generically predict that our universe today should have several distinct features—features that are currently being tested by the new generation of high-precision astronomical measurements. Even as inflation passes more and more stringent empirical tests, theorists continue to explore broader features and implications, such as what might have come before an inflationary epoch, how inflation might have ended within our observable universe, and how inflation might

Center for Theoretical Physics, Laboratory for Nuclear Science and Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

*To whom correspondence should be addressed. E-mail: guth@ctp.mit.edu (A.H.G.); dikaiser@mit.edu

the experimental frontier, tireless efforts are underway to detect CDM directly in underground low-background experiments; to produce CDM in particle accelerators; to make measurements to determine the nature of the dark energy (its sound speed, equation of state, and evolution); and to test inflation and determine what happened in the earliest moments of our universe.

References and Notes

1. A. Einstein, *Ann. der Phys.* **18**, 639 (1905).
2. A. S. Eddington, Address to the Mathematical and Physical Science Section (Section A) of the Royal Astronomical Society in Cardiff, and *Nature* **106**, 14 (1920).
3. H. A. Bethe, *Phys. Rev.* **55**, 103 (1939).
4. W. A. Fowler, G. R. Burbidge, E. M. Burbidge, *Astrophys. J.* **122**, 271 (1955).
5. A. Einstein, *Ann. der Phys.* **17**, 132 (1905).
6. A. Einstein, *Ann. der Phys.* **19**, 289 (1906).
7. A. Einstein, *Phys. Z.* **18**, 121 (1917).
8. A. Einstein, *Ann. der Phys.* **17**, 549 (1905).
9. A. Einstein, *Ann. der Phys.* **17**, 891 (1905).
10. A. Einstein, *Ann. der Phys.* **49**, 769 (1916).
11. Letter from Isaac Newton to Robert Hooke, 5 February 1675.
12. E. B. Fomalont, R. A. Sramek, *Phys. Rev. Lett.* **36**, 1475 (1976).
13. I. I. Shapiro, *Phys. Rev. Lett.* **13**, 789 (1964).
14. B. Bertotti, L. less, P. Tortora, *Nature* **425**, 374 (2003).
15. F. Zwicky, *Phys. Rev.* **51**, 290 (1937).
16. F. Zwicky, *Phys. Rev.* **51**, 679 (1937).
17. V. C. Rubin, N. Thonnard, W. K. Ford Jr., *Astrophys. J.* **225**, L107 (1978).
18. R. V. Pound, G. A. Rebka Jr., *Phys. Rev. Lett.* **4**, 337 (1960).
19. R. F. C. Vessot *et al.*, *Phys. Rev. Lett.* **45**, 2081 (1980).
20. A pulsar is a rapidly rotating neutron star that emits highly directional radiation. Pulses are observed as the rotation periodically sweeps the beam across the sky toward the observer, similar to a lighthouse.
21. J. H. Taylor, R. A. Hulse, L. A. Fowler, G. E. Gullahorn, J. M. Rankin, *Astrophys. J.* **206**, L53 (1976).
22. J. H. Taylor, L. A. Fowler, J. M. Weisberg, *Nature* **277**, 437 (1979).
23. M. Begelman, *Science* **300**, 1898 (2003).
24. L. Stella, M. Vietri, *Astrophys. J.* **492**, L59 (1998).
25. I. Ciufolini, E. C. Pavlis, *Nature* **431**, 958 (2004).
26. G. Gamow, *Phys. Rev.* **70**, 572 (1946).
27. R. A. Alpher, H. Bethe, G. Gamow, *Phys. Rev.* **73**, 803 (1948).
28. H. A. Bethe, *Phys. Rev.* **55**, 103 (1939).
29. S. Perlmutter *et al.*, *Astrophys. J.* **517**, 565 (1999).
30. A. G. Riess *et al.*, *Astron. J.* **116**, 1009 (1998).
31. R. A. Alpher, R. C. Herman, *Phys. Rev.* **75**, 1089 (1949).
32. J. C. Mather, D. J. Fixsen, R. A. Shafer, C. Mosier, D. T. Wilkinson, *Astrophys. J.* **512**, 511 (1999).
33. J. R. Bond, G. Efstathiou, *Astrophys. J.* **285**, L45 (1984).
34. For examples of balloon-borne CMB experiments, see the following Web sites: Far Infra-Red Survey (FIRS, http://physics.princeton.edu/~cmb/firs.html); Balloon Observations Of Millimetric Extragalactic Radiation and Geophysics (BOOMERanG, http://cmb.phys.cwru.edu/boomerang/); Millimeter Anisotropy eXperiment Imaging Array (MAXIMA, http://cosmology.berkeley.edu/group/cmb/index.html); QMAP (http://physics.princeton.edu/~cmb/qmap/qmap.html); Medium Scale Anisotropy Measurement and TopHat (MSAM/TopHat, http://topweb.gsfc.nasa.gov/); Archeops (www.archeops.org/index_english.html); Balloon-borne Anisotropy Measurement (BAM, http://cmbr.physics.ubc.ca/bam/experimental.html); ARGO (http://oberon.roma1.infn.it/argo/); Background Emission Anisotropy Scanning Telescope (BEAST/ACE, www.deepspace.ucsb.edu/research/Sphome.htm).
35. For examples of ground-based CMB experiments, see the following Web sites: Arcminute Cosmology Bolometer Array Receiver (ACBAR, http://cosmology.berkeley.edu/group/swlh/acbar/); Advanced Cosmic Microwave Explorer (ACME/HACME, www.deepspace.ucsb.edu/research/Sphome.htm); Antarctic Plateau Anisotropy CHasing Experiment (APACHE, www.bo.iasf.cnr.it/~valenziano/APACHE/apache.htm); Australia Telescope Compact Array (ATCA, www.atnf.csiro.au/research/cmbr/index.html); Cosmic Anisotropy Polarization MAPper (CAPMAP, http://wwwphy.princeton.edu/cosmology/); Cosmic Anisotropy Telescope (CAT, www.mrao.cam.ac.uk/telescopes/cat/index.html); Cosmic Background Imager (CBI, www.astro.caltech.edu/~tjp/CBI/); CosmicMicrowave Polarization at SmallScales (COMPASS, http://cmb.physics.wisc.edu/compass.html); Degree Angular Scale Interferometer (DASI, http://astro.uchicago.edu/dasi/); Princeton I, Q, and U Experiment (PIQUE, http://wwwphy.princeton.edu/cosmology/); Polarization Observations of Large Angular Regions (POLAR, http://cmb.physics.wisc.edu/polar/); Saskatoon/Toco/Mobile Anisotropy Telescope (SK/TOCO/MAT, http://wwwphy.princeton.edu/cosmology/); Python (http://astro.uchicago.edu/cara/research/cmbr/python.html); Tenerife (www.iac.es/project/cmb/rad/index.html); Very Small Array (VSA, www.mrao.cam.ac.uk/telescopes/vsa/index.html); Jodrell Bank; Viper; Owens Valley Radio Observatory (OVRO, www.astro.caltech.edu/~tjp/OVRO-CMB/intrinsic.html).
36. The Legacy Archive for Microwave Background Data Analysis (LAMBDA) data center, available at http://lambda.gsfc.nasa.gov, has links to data and further information about the CMB experiments.
37. C. L. Bennett *et al.*, *Astrophys. J.* **148**, 1 (2003).
38. W. L. Freedman *et al.*, *Astrophys. J.* **553**, 47 (2001).
39. W. J. Percival *et al.*, *Mon. Not. R. Astron. Soc.* **337**, 1068 (2002).
40. M. Tegmark *et al.*, *Phys. Rev.* **D69**, 103501 (2004).
41. P. J. E. Peebles, B. Ratra, *Rev. Mod. Phys.* **75**, 559 (2003).
42. P. Steinhardt, *Philos. Trans. Math. Phys. Eng. Sci.* **361**, 2497 (2003).
43. *Connecting Quarks with the Cosmos: Eleven Science Questions for the New Century* (National Academies Press, Washington, DC, 2003).

44. I thank R. Weiss, E. L. Wright, and G. Hinshaw for helpful comments that led to an improved paper. I am also grateful to B. Griswold for expert graphics assistance. Support for this work was provided by NASA.

10.1126/science.1106444

REVIEW

Inflationary Cosmology: Exploring the Universe from the Smallest to the Largest Scales

Alan H. Guth* and David I. Kaiser*

Understanding the behavior of the universe at large depends critically on insights about the smallest units of matter and their fundamental interactions. Inflationary cosmology is a highly successful framework for exploring these interconnections between particle physics and gravitation. Inflation makes several predictions about the present state of the universe—such as its overall shape, large-scale smoothness, and smaller scale structure—which are being tested to unprecedented accuracy by a new generation of astronomical measurements. The agreement between these predictions and the latest observations is extremely promising. Meanwhile, physicists are busy trying to understand inflation's ultimate implications for the nature of matter, energy, and spacetime.

The scientific community is celebrating the International Year of Physics in 2005, honoring the centennial of Albert Einstein's most important year of scientific innovation. In the span of just a few months during 1905 Einstein introduced key notions that would dramatically change our understanding of matter and energy as well as the nature of space and time. The centennial of these seminal developments offers an enticing opportunity to take stock of how scientists think about these issues today. We focus in particular on recent developments in the field of inflationary cosmology, which draws on a blend of concepts from particle physics and gravitation. The last few years have been a remarkably exciting time for cosmology, with new observations of unprecedented accuracy yielding many surprises. Einstein's legacy is flourishing in the early 21st century.

Inflation was invented a quarter of a century ago, and has become a central ingredient of current cosmological research. Describing dramatic events in the earliest history of our universe, inflationary models generically predict that our universe today should have several distinct features—features that are currently being tested by the new generation of high-precision astronomical measurements. Even as inflation passes more and more stringent empirical tests, theorists continue to explore broader features and implications, such as what might have come before an inflationary epoch, how inflation might have ended within our observable universe, and how inflation might

Center for Theoretical Physics, Laboratory for Nuclear Science and Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

*To whom correspondence should be addressed. E-mail: guth@ctp.mit.edu (A.H.G.); dikaiser@mit.edu

arise in the context of our latest understanding of the structure of space, time, and matter.

Particle theory has been changing rapidly, and these theoretical developments have provided just as important a spur to inflationary cosmology as have the new observations. During the 1960s and 70s, particle physicists discovered that if they neglected gravity, they could construct highly successful descriptions of three out of the four basic forces in the universe: electromagnetism and the strong and weak nuclear forces. The "standard model of particle physics," describing these three forces, was formulated within the framework of quantum field theory, the physicist's quantum-mechanical description of subatomic matter. Inflationary cosmology was likewise first formulated in terms of quantum field theory. Now, however, despite (or perhaps because of) the spectacular experimental success of the standard model, the major thrust of particle physics research is aimed at moving beyond it.

For all its successes, the standard model says nothing at all about the fourth force: gravity. For more than 50 years physicists have sought ways to incorporate gravity within a quantum-mechanical framework, initially with no success. But for the past 25 or more years, an ever-growing group of theoretical physicists has been pursuing superstring theory as the bright hope for solving this problem. To accomplish this task, however, string theorists have been forced to introduce many novel departures from conventional ideas about fundamental forces and the nature of the universe. For one thing, string theory stipulates that the basic units of matter are not pointlike particles (as treated by quantum field theory), but rather one-dimensional extended objects, or strings. Moreover, in order to be mathematically self-consistent, string theories require the existence of several additional spatial dimensions. Whereas our observable universe seems to contain one timelike dimension and three spatial dimensions—height, width, and depth—string theory postulates that our universe actually contains at least six additional spatial dimensions, each at right angles to the others and yet somehow hidden from view.

For measurements at low energies, string theory should behave effectively like a quantum field theory, reproducing the successes of the standard model of particle physics. Yet the interface between cosmology and string theory has been a lively frontier. For example, some theorists have been constructing inflationary models for our universe that make use of the extra dimensions that string theory introduces. Others have been studying the string theory underpinnings for inflationary models, exploring such topics as the nature of vacuum states and the question of their uniqueness. As we will see, inflation continues to occupy a central place in cosmological research, even as its relation to fundamental particle physics continues to evolve.

Inflationary Basics

According to inflationary cosmology (*1*–*3*), the universe expanded exponentially quickly for a fraction of a second very early in its history—growing from a patch as small as 10^{-26} m, one hundred billion times smaller than a proton, to macroscopic scales on the order of a meter, all within about 10^{-35} s—before slowing down to the more stately rate of expansion that has characterized the universe's behavior ever since. The driving force behind this dramatic growth, strangely enough, was gravity. [For technical introductions to inflationary cosmology, see (*4*–*6*); a more popular description may be found in (*7*).] Although this might sound like hopeless (or, depending on one's inclinations, interesting) speculation, in fact inflationary cosmology leads to several quantitative predictions about the present behavior of our universe—predictions that are being tested to unprecedented accuracy by a new generation of observational techniques. So far the agreement has been excellent.

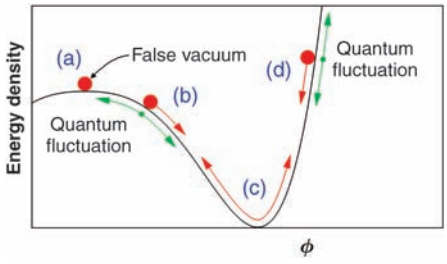


Fig. 1. In simple inflationary models, the universe at early times is dominated by the potential energy density of a scalar field, ϕ. Red arrows show the classical motion of ϕ. When ϕ is near region (a), the energy density will remain nearly constant, $\rho \cong \rho_f$, even as the universe expands. Moreover, cosmic expansion acts like a frictional drag, slowing the motion of ϕ: Even near regions (b) and (d), ϕ behaves more like a marble moving in a bowl of molasses, slowly creeping down the side of its potential, rather than like a marble sliding down the inside of a polished bowl. During this period of "slow roll," ρ remains nearly constant. Only after ϕ has slid most of the way down its potential will it begin to oscillate around its minimum, as in region (c), ending inflation.

How could gravity drive the universal repulsion during inflation? The key to this rapid expansion is that in Einstein's general relativity (physicists' reigning description of gravity), the gravitational field couples both to mass-energy (where mass and energy are interchangeable thanks to Einstein's $E = mc^2$) and to pressure, rather than to mass alone. In the simplest scenario, in which at least the observable portion of our universe can be approximated as being homogeneous and isotropic—that is, having no preferred locations or directions—Einstein's gravitational equations give a particularly simple result. The expansion of the universe may be described by introducing a time-dependent "scale factor," $a(t)$, with the separation between any two objects in the universe being proportional to $a(t)$. Einstein's equations prescribe how this scale factor will evolve over time, t. The rate of acceleration is proportional to the density of mass-energy in the universe, ρ, plus three times its pressure, p: $d^2a/dt^2 = -4\pi G\ (\rho + 3p)a/3$, where G is Newton's gravitational constant (and we use units for which the speed of light $c = 1$). The minus sign is important: Ordinary matter under ordinary circumstances has both positive mass-energy density and positive (or zero) pressure, so that $(\rho + 3p) > 0$. In this case, gravity acts as we would expect it to: All of the matter in the universe tends to attract all of the other matter, causing the expansion of the universe as a whole to slow down.

The crucial idea behind inflation is that matter can behave rather differently from this familiar pattern. Ideas from particle physics suggest that the universe is permeated by scalar fields, such as the Higgs field of the standard model of particle physics, or its more exotic generalizations. (A scalar field takes exactly one value at every point in space and time. For example, one could measure the temperature at every position in a room and repeat the measurements over time, and represent the measurements by a scalar field, T, of temperature. Electric and magnetic fields are vector fields, which carry three distinct components at every point in space and time: the field in the x direction, in the y direction, and in the z direction. Scalar fields are introduced in particle physics to describe certain kinds of particles, just as photons are described in quantum field theories in terms of electromagnetic fields.) These scalar fields can exist in a special state, having a high energy density that cannot be rapidly lowered, such as the arrangement labeled (a) in Fig. 1. Such a state is called a "false vacuum." Particle physicists use the word "vacuum" to denote the state of lowest energy. "False vacua" are only metastable, not the true states of lowest possible energy.

In the early universe, a scalar field in such a false vacuum state can dominate all the contributions to the total mass-energy density, ρ. During this period, ρ remains nearly constant, even as the volume of the universe expands rapidly: $\rho \cong \rho_f$ = constant. This is quite different from the density of ordinary matter, which decreases when the volume of its container increases. Moreover, the first law of thermodynamics, in the context of general relativity, implies that if $\rho \cong \rho_f$ while the universe expands, then the equation of state for this special state of matter must be $p \cong -\rho_f$, a negative pressure. This yields $d^2a/dt^2 = 8\pi G\rho_f\ a/3$: Rather than slowing down, the cosmic expansion rate will grow rapidly, driven by the neg-

ative pressure created by this special state of matter. Under these circumstances, the scale factor grows as $a \propto e^{Ht}$, where the Hubble parameter, $H \equiv a^{-1}da/dt$, which measures the universe's rate of expansion, assumes the constant value, $H \cong (8\pi G\rho_f/3)^{1/2}$. The universe expands exponentially until the scalar field rolls to near the bottom of the hill in the potential energy diagram.

What supplies the energy for this gigantic expansion? The answer, surprisingly, is that no energy is needed (*7*). Physicists have known since the 1930s (*8*) that the gravitational field carries *negative* potential energy density. As vast quantities of matter are produced during inflation, a vast amount of negative potential energy materializes in the gravitational field that fills the ever-enlarging region of space. The total energy remains constant, and very small, and possibly exactly equal to zero.

There are now dozens of models that lead to this generic inflationary behavior, featuring an equation of state, $p \cong -\rho$, during the early universe (*4*, *9*). This entire family of models, moreover, leads to several main predictions about today's universe. First, our observable universe should be spatially flat. Einstein's general relativity allows for all kinds of curved (or "non-Euclidean") spacetimes. Homogeneous and isotropic spacetimes fall into three classes (Fig. 2), depending on the value of the mass-energy density, ρ. If $\rho > \rho_c$, where $\rho_c \equiv 3H^2/(8\pi G)$, then Einstein's equations imply that the spacetime will be positively curved, or closed (akin to the two-dimensional surface of a sphere); parallel lines will intersect, and the interior angles of a triangle will always add up to more than 180°. If $\rho < \rho_c$, the spacetime will be negatively curved, or open (akin to the two-dimensional surface of a saddle); parallel lines will diverge and triangles will sum to less than 180°. Only if $\rho = \rho_c$ will spacetime be spatially flat (akin to an ordinary two-dimensional flat surface); in this case, all of the usual rules of Euclidean geometry apply. Cosmologists use the letter Ω to designate the ratio of the actual mass-energy density in the universe to this critical value: $\Omega \equiv \rho/\rho_c$. Although general relativity allows any value for this ratio, inflation predicts that $\Omega = 1$ within our observable universe to extremely high accuracy. Until recently, uncertainties in the measurement of Ω allowed any value in the wide range, $0.1 \leq \Omega \leq 2$, with many observations seeming to favor $\Omega \cong 0.3$. A new generation of detectors, however, has dramatically changed the situation. The latest observations, combining data from the Wilkinson Microwave Anisotropy Probe (WMAP), the Sloan Digital Sky Survey (SDSS), and observations of type Ia supernovae, have measured $\Omega = 1.012^{+0.018}_{-0.022}$ (*10*)—an amazing match between prediction and observation.

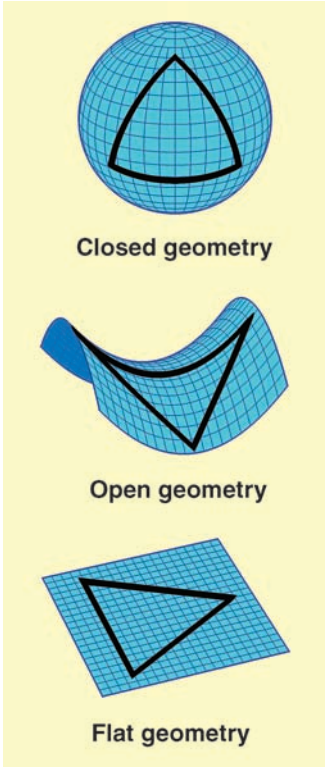


Fig. 2. According to general relativity, spacetime may be warped or curved, depending on the density of mass-energy. Inflation predicts that our observable universe should be spatially flat to very high accuracy.

In fact, inflation offers a simple explanation for why the universe should be so flat today. In the standard big bang cosmology (without inflation), $\Omega = 1$ is an *unstable* solution: If Ω were ever-so-slightly less than 1 at an early time, then it would rapidly slide toward 0. For example, if Ω were 0.9 at 1 s after the big bang, it would be only 10^{-14} today. If Ω were 1.1 at $t =$ 1 s, then it would have grown so quickly that the universe would have recollapsed just 45 s later. In inflationary models, on the other hand, any original curvature of the early universe would have been stretched out to near-flatness as the universe underwent its rapid expansion (Fig. 3). Quantitatively, $|\Omega - 1| \propto 1/(aH)^2$, so that while $H \cong$ constant and $a \propto e^{Ht}$ during the inflationary epoch, Ω gets driven rapidly to 1.

The second main prediction of inflation is that the presently observed universe should be remarkably smooth and homogeneous on the largest astronomical scales. This, too, has been measured to extraordinary accuracy during the past decades. Starting in the 1960s, Earth-bound, balloon-borne, and now satellite detectors have measured the cosmic microwave background (CMB) radiation, a thermal bath of photons that fills the sky. The photons today have a frequency that corresponds to a temperature of 2.728 K (*11*). These photons were released ~400,000 years after the inflationary epoch, when the universe had cooled to a low enough temperature that would allow stable (and electrically neutral) atoms to form. Before that time, the ambient temperature of matter in the universe was so high that would-be atoms were broken up by high-energy photons as soon as they formed, so that the photons were effectively trapped, constantly colliding into electrically charged matter. Since stable atoms formed, however, the CMB photons have been streaming freely. Their temperature today is terrifically uniform: After adjusting for the Earth's motion, CMB photons measured from any direction in the sky have the same temperature to one part in 10^5 (*12*).

Without inflation, this large-scale smoothness appears quite puzzling. According to ordinary (noninflationary) big bang cosmology, these photons should never have had a chance to come to thermal equilibrium: The regions in the sky from which they were released would have been about 100 times farther apart than even light could have traveled between the time of the big bang and the time of the photons' release (*1*, *4*–*6*). Much like the flatness problem, inflation provides a simple and generic reason for the observed homogeneity of the CMB: Today's observable universe originated from a *much* smaller region than that in the noninflationary scenarios. This much-smaller patch could easily have become smooth before inflation began. Inflation would then stretch this small homogeneous region to encompass the entire observable universe.

A third major prediction of inflationary cosmology is that there should be tiny departures from this strict large-scale smoothness and that these ripples (or "perturbations") should have a characteristic spectrum. Today these ripples can be seen directly as fluctuations in the CMB. Although the ripples are believed to be responsible for the grandest structures of the universe—galaxies, superclusters, and giant voids—in inflationary models they arise from quantum fluctuations, usually important only on atomic scales or smaller. The field ϕ that drives inflation, like all quantum fields, undergoes quantum fluctuations in accord with the Heisenberg uncertainty principle. During inflation these quantum fluctuations are stretched proportionally to $a(t)$, rapidly growing to macroscopic scales. The result: a set of nearly scale-invariant perturbations extending over a huge range of wavelengths (*13*). Cosmologists parameterize the spectrum of primordial perturbations by a spectral index, n_s. A scale-invariant spectrum would have $n_s = 1.00$; inflationary models generically predict $n_s = 1$ to within ~10%. The latest measurements of these perturbations by WMAP and SDSS reveal $n_s = 0.977^{+0.039}_{-0.025}$ (*10*).

Until recently, astronomers were aware of several cosmological models that were consistent with the known data: an open universe, with $\Omega \cong 0.3$; an inflationary universe

with considerable dark energy (Λ); an inflationary universe without Λ; and a universe in which the primordial perturbations arose from topological defects such as cosmic strings. Dark energy (*14*) is a form of matter with negative pressure that is currently believed to contribute about 70% of the total energy of the observed universe. Cosmic strings are long, narrow filaments hypothesized to be scattered throughout space, remnants of a symmetry-breaking phase transition in the early universe (*15*, *16*). [Cosmic strings are topologically nontrivial configurations of fields, which should not be confused with the fundamental strings of superstring theory. The latter are usually believed to have lengths on the order of 10^{-35} m, although for some compactifications these strings might also have astronomical lengths (*17*).] Each of these models leads to a distinctive pattern of resonant oscillations in the early universe, which can be probed today through its imprint on the CMB. As can be seen in Fig. 4 (*18*), three of the models are now definitively ruled out. The full class of inflationary models can make a variety of predictions, but the prediction of the simplest inflationary models with large Λ, shown on the graph, fits the data beautifully.

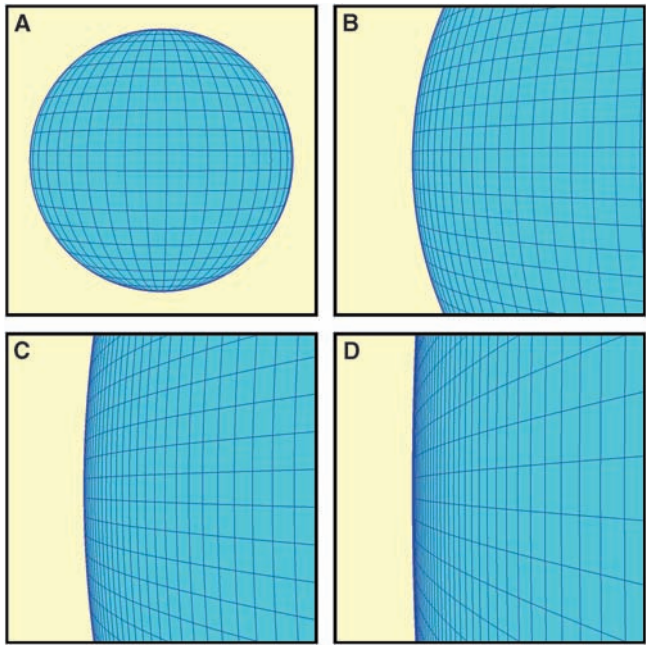


Fig. 3. (**A** to **D**) The expanding sphere illustrates the solution to the flatness problem in inflationary cosmology. As the sphere becomes larger, its surface becomes more and more flat. In the same way, the exponential expansion of spacetime during inflation causes it to become spatially flat.

Before and After Inflation

Research in recent years has included investigations of what might have preceded inflation and how an inflationary epoch might have ended.

Soon after the first inflationary models were introduced, several physicists (*19*–*21*) realized that once inflation began, it would in all likelihood never stop. Regions of space would stop inflating, forming what can be called "pocket universes," one of which would contain the observed universe. Nonetheless, at any given moment *some* portion of the universe would still be undergoing exponential expansion, in a process called "eternal inflation." In the model depicted in Fig. 1, for example, quantum-mechanical effects compete with the classical motion to produce eternal inflation. Consider a region of size H^{-1}, in which the average value of ϕ is near (b) or (d) in the diagram. Call the average energy density ρ_0. Whereas the classical tendency of ϕ is to roll slowly downward (red arrow) toward the minimum of its potential, the field will also be subject to quantum fluctuations (green arrows) similar to those described above. The quantum fluctuations will give the field a certain likelihood of hopping up the wall of potential energy rather than down it. Over a time period H^{-1}, this region will grow $e^3 \cong 20$ times its original size. If the probability that the field will roll up the potential hill during this period is greater than 1/20, then on average the volume of space in which $\rho > \rho_0$ increases with time (*4*, *21*, *22*). The probability of upward fluctuations tends to become large when the initial value of ϕ is near the peak at (a) or high on the hill near (d), so for most potential energy functions the condition for eternal inflation is attainable. In that case the volume of the inflating region grows exponentially, and forever: Inflation would produce an infinity of pocket universes.

An interesting question is whether or not eternal inflation makes the big bang unnecessary: Might eternal inflation have been truly "eternal," existing more or less the same way for all time, or is it only "eternal" to the future once it gets started? Borde and Vilenkin have analyzed this question (most recently, with Guth), and have concluded that eternal inflation could not have been past-eternal: Using kinematic arguments, they showed (*23*) that the inflating region must have had a past boundary, before which some alternative description must have applied. One possibility would be the creation of the universe by some kind of quantum process.

Another major area of research centers on the mechanisms by which inflation might have ended within our observable universe. The means by which inflation ends have major consequences for the subsequent history of our universe. For one thing, the colossal expansion during inflation causes the temperature of the universe to plummet nearly to zero, and dilutes the density of ordinary matter to negligible quantities. Some mechanism must therefore convert the energy of the scalar field, ϕ, into a hot soup of garden-variety matter.

In most models, inflation ends when ϕ oscillates around the minimum of its potential, as in region (c) of Fig. 1. Quantum-mechanically, these field oscillations correspond to a collection of ϕ particles approximately at rest. Early studies of postinflation "reheating" assumed that individual ϕ particles would decay during these oscillations like radioactive nuclei. More recently, it has been discovered (*24*–*28*) that these oscillations would drive resonances in ϕ's interactions with other quantum fields. Instead of individual ϕ particles decaying independently, these resonances would set up collective behavior—ϕ would release its energy more like a laser than an ordinary light bulb, pouring it extremely rapidly into a sea of newly created particles. Large numbers of particles would be created very quickly within specific energy-bands, corresponding to the frequency of ϕ's oscillations and its higher harmonics.

This dramatic burst of particle creation would affect spacetime itself, as it responded to changes in the arrangement of matter and energy. The rapid transfer of energy would excite gravitational perturbations, of which the most strongly amplified would be those with frequencies within the resonance bands of the decaying ϕ field. In some extreme cases, very long-wavelength perturbations can be amplified during reheating, which could in principle even leave an imprint on the CMB (*29*).

Brane Cosmology: Sticking Close to Home

Although superstring theory promises to synthesize general relativity with the other fundamental forces of nature, it introduces a number of surprising features—such as the existence of microscopic strings, rather than particles, as the fundamental units of matter, along with the existence of several extra spatial dimensions in the universe. Could our observable universe really be built from such a bizarre collection of ingredients?

Naïvely, one might expect the extra dimensions to conflict with the observed behavior of gravity. To be successful, string theory, like general relativity, must reduce to Newton's law of gravity in the appropriate limit. In

Newton's formulation, gravity can be described by force lines that always begin and end on masses. If the force lines could spread in n spatial dimensions, then at a radius r from the center, they would intersect a hypersphere with surface area proportional to r^{n-1}. An equal number of force lines would cross the hypersphere at each radius, which means that the density of force lines would be proportional to $1/r^{n-1}$. For $n = 3$, this reproduces the familiar Newtonian force law, $F \propto 1/r^2$, which has been tested (along with its Einsteinian generalization) to remarkable accuracy over a huge range of distances, from astronomical scales down to less than a millimeter (*30*, *31*).

An early response to this difficulty was to assume that the extra spatial dimensions are curled up into tiny closed circles rather than extending to macroscopic distances. Because gravity has a natural scale, known as the Planck length, $l_P \equiv (\hbar G/c^3)^{1/2} \cong 10^{-35}$ m (where $\hbar$ is Planck's constant divided by 2π), physicists assumed that l_P sets the scale for these extra dimensions. Just as the surface of a soda straw would appear one-dimensional when viewed from a large distance—even though it is really two-dimensional—our space would appear three-dimensional if the extra dimensions were "compactified" in this way. On scales much larger than the radii of the extra dimensions, r_c, we would fail to notice them: The strength of gravity would fall off in its usual $1/r^2$ manner for distances $r \gg r_c$, but would fall off as $1/r^{n-1}$ for scales $r \ll r_c$ (*32*). The question remained, however, what caused this compactification, and why this special behavior affected only some but not all dimensions.

Recently, Arkani-Hamed, Dimopoulos, and Dvali (*33*) realized that there is no necessary relation between ℓ_P and r_c, and that experiments only require $r_c \leq 1$ mm. Shortly afterward, Randall and Sundrum (*34*, *35*) discovered that the extra dimensions could even be infinite in extent! In the Randall-Sundrum model, our observable universe lies on a membrane, or "brane" for short, of three spatial dimensions, embedded within some larger multidimensional space. The key insight is that the energy carried by the brane will sharply affect the way the gravitational field behaves. For certain spacetime configurations, the behavior of gravity along the brane can appear four-dimensional (three space and one time), even in the presence of extra dimensions. Gravitational force lines would tend to "hug" the brane, rather than spill out into the "bulk"—the spatial volume in which our brane is embedded. Along the brane, therefore, the dominant behavior of the gravitational force would still be $1/r^2$.

In simple models, in which the spacetime geometry along our brane is highly symmetric, such as the Minkowski spacetime of special relativity, the effective gravitational field along our brane is found to mimic the usual Einsteinian results to high accuracy (*36*, *37*). At very short distances there are calculable (and testable) deviations from standard gravity, and there may also be deviations for very strong gravitational fields, such as those near black holes. There are also modifications to the cosmological predictions of gravity. In the usual case, when Einstein's equations are applied to a homogeneous and isotropic spacetime, one finds $H^2 \propto \rho - k/a^2$, where k is a constant connected to the curvature of the universe. If instead we lived on a brane embedded within one large extra dimension, then $H^2 \propto \rho + \alpha\rho^2 - k/a^2$, where α is a constant (*38*).

Under ordinary conditions, ρ decreases as the universe expands, and so the new term in the effective Einstein equations should have minimal effects at late times in our observable universe. But we saw above that during an inflationary epoch, $\rho \cong$ constant, and in these early moments the departures from the ordinary Einsteinian case can be dramatic. In particular, the ρ^2 term would allow inflation to occur at lower energies than are usually assumed in ordinary (nonembedded) models, with potential energy functions that are less flat than are ordinarily needed to sustain inflation. Moreover, the spectrum of primordial perturbations would get driven even closer to the scale-invariant shape, with $n_s = 1.00$ (*39*, *40*). Brane cosmology thus leads to some interesting effects during the early universe, making inflation even more robust than in ordinary scenarios.

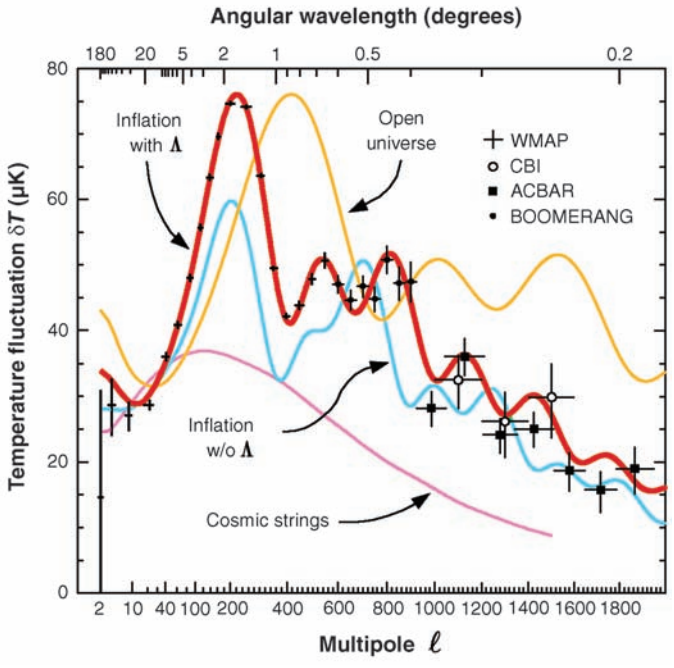


Fig. 4. Comparison of the latest observational measurements of the temperature fluctuations in the CMB with several theoretical models, as described in the text. The temperature pattern on the sky is expanded in multipoles (i.e., spherical harmonics), and the intensity is plotted as a function of the multipole number ℓ. Roughly speaking, each multipole ℓ corresponds to ripples with an angular wavelength of $360°/\ell$.

String Cosmology

Although theories of extra dimensions establish a connection between string theory and cosmology, the developments of the past few years have pushed the connection much further. [For reviews, see (*41*–*43*).] The union of string theory and cosmology is barely past its honeymoon, but so far the marriage appears to be a happy one. Inflation, from its inception, was a phenomenologically very successful idea that has been in need of a fundamental theory to constrain its many variations. String theory, from its inception, has been a very well-constrained mathematical theory in need of a phenomenology to provide contact with observation. The match seems perfect, but time will be needed before we know for sure whether either marriage partner can fulfill the needs of the other. In the meantime, ideas are stirring that have the potential to radically alter our ideas about fundamental laws of physics.

For many years the possibility of describing inflation in terms of string theory seemed completely intractable, because the only string vacua that were understood were highly supersymmetric ones, with many massless scalar fields, called moduli, which have potential energy functions that vanish identically to all orders of perturbation theory. When the effects of gravity are included, the energy density of such supersymmetric states is never positive. Inflation, on the other hand, requires a positive energy density, and it requires a hill in the potential energy function. Inflation, therefore, could only be contemplated in the context of nonperturbative supersymmetry-breaking effects, of which there was very little understanding.

The situation changed dramatically with the realization that string theory contains not only strings, but also branes, and fluxes,

which can be thought of as higher-dimensional generalizations of magnetic fields. The combination of these two ingredients makes it possible to construct string theory states that break supersymmetry and that give nontrivial potential energy functions to all the scalar fields.

One very attractive idea for incorporating inflation into string theory is to use the positions of branes to play the role of the scalar field that drives inflation. The earliest version of this theory was proposed in 1998 by Dvali and Tye (*44*), shortly after the possibility of large extra dimensions was proposed in (*33*). In the Dvali-Tye model, the observed universe is described not by a single three-dimensional brane, but instead by a number of three-dimensional branes, which in the vacuum state would sit on top of each other. If some of the branes were displaced, however, in a fourth spatial direction, then the energy would be increased. The brane separation would be a function of time and the three spatial coordinates along the branes, and so from the point of view of an observer on the brane, it would act like a scalar field that could drive inflation. At this stage, however, the authors needed to invoke unknown mechanisms to break supersymmetry and to give the moduli fields nonzero potential energy functions.

In 2003, Kachru, Kallosh, Linde, and Trivedi (*45*) showed how to construct complicated string theory states for which all the moduli have nontrivial potentials, for which the energy density is positive, and for which the approximations that were used in the calculations appeared justifiable. These states are only metastable, but their lifetimes can be vastly longer than the 14 billion years that have elapsed since the big bang. There was nothing elegant about this construction—the six extra dimensions implied by string theory are curled not into circles, but into complicated manifolds with a number of internal loops that can be threaded by several different types of flux, and populated by a hodgepodge of branes. Joined by Maldacena and McAllister, this group (*46*) went on to construct states that can describe inflation, in which a parameter corresponding to a brane position can roll down a hill in its potential energy diagram. Generically the potential energy function is not flat enough for successful inflation, but the authors argued that the number of possible constructions was so large that there may well be a large class of states for which sufficient inflation is achieved. Iizuka and Trivedi (*47*) showed that successful inflation can be attained by curling the extra dimensions into a manifold that has a special kind of symmetry.

A tantalizing feature of these models is that at the end of inflation, a network of strings would be produced (*17*). These could be fundamental strings, or branes with one spatial dimension. The CMB data of Fig. 4 rule out the possibility that these strings are major sources of density fluctuations, but they are still allowed if they are light enough so that they do not disturb the density fluctuations from inflation. String theorists are hoping that such strings may be able to provide an observational window on string physics.

A key feature of the constructions of inflating states or vacuumlike states in string theory is that they are far from unique. The number might be something like 10^{500} (*48–50*), forming what Susskind has dubbed the ''landscape of string theory.'' Although the rules of string theory are unique, the low-energy laws that describe the physics that we can in practice observe would depend strongly on which vacuum state our universe was built upon. Other vacuum states could give rise to different values of ''fundamental'' constants, or even to altogether different types of ''elementary'' particles, and even different numbers of large spatial dimensions! Furthermore, because inflation is generically eternal, one would expect that the resulting eternally inflating spacetime would sample every one of these states, each an infinite number of times. Because all of these states are possible, the important problem is to learn which states are probable. This problem involves comparison of one infinity with another, which is in general not a well-defined question (*51*). Proposals have been made and arguments have been given to justify them (*52*), but no conclusive solution to this problem has been found.

What, then, determined the vacuum state for our observable universe? Although many physicists (including the authors) hope that some principle can be found to understand how this choice was determined, there are no persuasive ideas about what form such a principle might take. It is possible that inflation helps to control the choice of state, because perhaps one state or a subset of states expands faster than any others. Because inflation is generically eternal, the state that inflates the fastest, along with the states that it decays into, might dominate over any others by an infinite amount. Progress in implementing this idea, however, has so far been nil, in part because we cannot identify the state that inflates the fastest, and in part because we cannot calculate probabilities in any case. If we could calculate the decay chain of the most rapidly inflating state, we would have no guarantee that the number of states with significant probability would be much smaller than the total number of possible states.

Another possibility, now widely discussed, is that *nothing* determines the choice of vacuum for our universe; instead, the observable universe is viewed as a tiny speck within a multiverse that contains *every* possible type of vacuum. If this point of view is right, then a quantity such as the electron-to-proton mass ratio would be on the same footing as the distance between our planet and the sun. Neither is fixed by the fundamental laws, but instead both are determined by historical accidents, restricted only by the fact that if these quantities did not lie within a suitable range, we would not be here to make the observations. This idea—that the laws of physics that we observe are determined not by fundamental principles, but instead by the requirement that intelligent life can exist to observe them—is often called the anthropic principle. Although in some contexts this principle might sound patently religious, the combination of inflationary cosmology and the landscape of string theory gives the anthropic principle a scientifically viable framework.

A key reason why the anthropic approach is gaining attention is the observed fact that the expansion of the universe today is accelerating, rather than slowing down under the influence of normal gravity. In the context of general relativity, this requires that the energy of the observable universe is dominated by dark energy. The identity of the dark energy is unknown, but the simplest possibility is that it is the energy density of the vacuum, which is equivalent to what Einstein called the cosmological constant. To particle physicists it is not surprising that the vacuum has nonzero energy density, because the vacuum is known to be a very complicated state, in which particle-antiparticle pairs are constantly materializing and disappearing, and fields such as the electromagnetic field are constantly undergoing wild fluctuations. From the standpoint of the particle physicist, the shocking thing is that the energy density of the vacuum is so low. No one really knows how to calculate the energy density of the vacuum, but naïve estimates lead to numbers that are about 10^{120} times larger than the observational upper limit. There are both positive and negative contributions, but physicists have been trying for decades to find some reason why the positive and negative contributions should cancel, so far to no avail. It seems even more hopeless to find a reason why the net energy density should be nonzero, but 120 orders of magnitude smaller than its expected value. However, if one adopts the anthropic point of view, it was argued as early as 1987 by Weinberg (*53*) that an explanation is at hand: If the multiverse contained regions with all conceivable values of the cosmological constant, galaxies and hence life could appear only in those

very rare regions where the value is small, because otherwise the huge gravitational repulsion would blow matter apart without allowing it to collect into galaxies.

The landscape of string theory and the evolution of the universe through the landscape are of course still not well understood, and some have argued (*54*) that the landscape might not even exist. It seems too early to draw any firm conclusions, but clearly the question of whether the laws of physics are uniquely determined, or whether they are environmental accidents, is an issue too fundamental to ignore.

Conclusions

During the past decade, cosmology has unquestionably entered the domain of high-precision science. Just a few years ago several basic cosmological quantities, such as the expansion parameter, H, and the flatness parameter, Ω, were known only to within a factor of 2. Now new observations using WMAP, SDSS, and the high-redshift type Ia supernovae measure these and other crucial quantities with percent-level accuracy. Several of inflation's most basic quantitative predictions, including $\Omega = 1$ and $n_s \cong 1$, may now be compared with data that are discriminating enough to distinguish inflation from many of its theoretical rivals. So far, every measure has been favorable to inflation.

Even with the evidence in favor of inflation now stronger than ever, much work remains. Inflationary cosmology has always been a framework for studying the interconnections between particle physics and gravitation—a collection of models rather than a unique theory. The next generation of astronomical detectors should be able to distinguish between competing inflationary models, whittling down the large number of options to a preferred few. One important goal is the high-precision measurement of polarization effects in the CMB, which allows the possibility of uncovering the traces of gravity waves originating from inflation. Gravity waves of the right pattern would be a striking test of inflation, and would allow us to determine the energy density of the ''false vacuum'' state that drove inflation. The new cosmological observations also offer physicists one of the best resources for evaluating the latest developments in idea-rich (but data-poor) particle theory, where much of the current research has been aimed at the high-energy frontier, well beyond the range of existing accelerators. Perhaps the interface between string theory and cosmology will lead to new predictions for the astronomers to test. Whether such tests are successful or not, physicists are certain to learn important lessons about the nature of space, time, and matter.

Meanwhile, several major puzzles persist. Now that physicists and astronomers are confident that $\Omega = 1$ to high accuracy, the question remains of just what type of matter and energy is filling the universe. Ordinary matter, such as the protons, neutrons, and electrons that make up atoms, contributes just 4% to this cosmic balance. Nearly one-quarter of the universe's mass-energy is some form of ''dark matter''—different in kind from the garden-variety matter we see around us, and yet exerting a measurable gravitational tug that shapes the way galaxies behave. Particle physicists have offered many candidates for this exotic dark matter, but to date no single contender has proved fully convincing (*55*).

Even more bizarre is the dark energy now known to contribute about 70% to Ω. This dark energy is driving a mini-inflationary epoch today, billions of years after the initial round of inflation. Today's accelerated expansion is far less fast than the earlier inflationary rate had been, but the question remains why it is happening at all. Could the dark energy be an example of Einstein's cosmological constant? Or maybe it is a variation on an inflationary theme: Perhaps some scalar field has been sliding down its potential energy hill on a time scale of billions of years rather than fractions of a second (*14*). Whatever its origin, dark energy, much like dark matter, presents a fascinating puzzle that will keep cosmologists busy for years to come.

References and Notes

1. A. H. Guth, *Phys. Rev. D* **23**, 347 (1981).
2. A. D. Linde, *Phys. Lett. B* **108**, 389 (1982).
3. A. Albrecht, P. J. Steinhardt, *Phys. Rev. Lett.* **48**, 1220 (1982).
4. A. D. Linde, *Particle Physics and Inflationary Cosmology* (Harwood, Philadelphia, 1990).
5. E. W. Kolb, M. S. Turner, *The Early Universe* (Addison-Wesley, Reading, MA, 1990).
6. A. R. Liddle, D. H. Lyth, *Cosmological Inflation and Large-Scale Structure* (Cambridge Univ. Press, New York, 2000).
7. A. H. Guth, *The Inflationary Universe: The Quest for a New Theory of Cosmic Origins* (Addison-Wesley, Reading, MA, 1997).
8. R. C. Tolman, *Phys. Rev.* **39**, 320 (1932).
9. D. H. Lyth, A. Riotto, *Phys. Rep.* **314**, 1 (1999).
10. M. Tegmark *et al.*, *Phys. Rev. D* **69**, 103501 (2004).
11. D. J. Fixsen *et al.*, *Astrophys. J.* **473**, 576 (1996).
12. D. N. Spergel *et al.*, *Astrophys. J. Suppl.* **148**, 175 (2003).
13. For a review, see V. F. Mukhanov, H. A. Feldman, R. H. Brandenberger, *Phys. Rep.* **215**, 203 (1992) and also (*6*).
14. For a review, see R. P. Kirshner, *Science* **300**, 1914 (2003).
15. A. Vilenkin, E. P. S. Shellard, *Cosmic Strings and other Topological Defects* (Cambridge Univ. Press, New York, 1994).
16. U.-L. Pen, U. Seljak, N. Turok, *Phys. Rev. Lett.* **79**, 1611 (1997).
17. For a review, see J. Polchinski, http://arxiv.org/abs/hep-th/0412244.
18. We thank M. Tegmark for providing this graph, which shows the most precise data points for each range of ℓ from recent observations, as summarized in (*10*). The cosmic string prediction is taken from (*16*). The other curves were all calculated for $n_s = 1$, $\Omega_{baryon} = 0.05$, and $H = 70$ km s^{-1} Mpc^{-1}, with the remaining parameters fixed as follows. "Inflation with Λ": Ω_{DM} (dark matter) = 0.23, $\Omega_\Lambda = 0.72$, and optical depth parameter $\tau = 0.17$; "Inflation without Λ": $\Omega_{DM} = 0.95$, $\Omega_\Lambda = 0$, $\tau = 0.06$; "Open universe": $\Omega_{DM} = 0.25$, $\Omega_\Lambda = 0$, $\tau = 0.06$. The 1-SD error bars include both observational uncertainty and "cosmic variance," the intrinsic quantum uncertainty in the predictions, as calculated from the "Inflation with Λ" model. With our current ignorance of the underlying physics, none of these theories predicts the overall amplitude of the fluctuations; the "Inflation with Λ" curve was normalized for a best fit, and the others were normalized arbitrarily.
19. P. J. Steinhardt, in *The Very Early Universe*, G. W. Gibbons, S. W. Hawking, S. T. C. Siklos, Eds. (Cambridge Univ. Press, New York, 1983), p. 251.
20. A. Vilenkin, *Phys. Rev. D* **27**, 2848 (1983).
21. A. D. Linde, *Phys. Lett. B* **175**, 395 (1986).
22. For a review, see A. H. Guth, *Phys. Rep.* **333-334**, 555 (2000).
23. A. Borde, A. H. Guth, A. Vilenkin, *Phys. Rev. Lett.* **90**, 151301 (2003).
24. L. A. Kofman, A. D. Linde, A. A. Starobinsky, *Phys. Rev. Lett.* **73**, 3195 (1994).
25. Y. Shtanov, J. H. Traschen, R. H. Brandenberger, *Phys. Rev. D* **51**, 5438 (1995).
26. D. Boyanovsky *et al.*, *Phys. Rev. D* **51**, 4419 (1995).
27. D. I. Kaiser, *Phys. Rev. D* **53**, 1776 (1996).
28. For a review, see L. A. Kofman, in *COSMO-97: Proceedings*, L. Roszkowski, Ed. (World Scientific, Singapore, 1998), pp. 312–321.
29. B. A. Bassett, D. I. Kaiser, R. Maartens, *Phys. Lett. B* **455**, 84 (1999).
30. Clifford Will, *Theory and Experiment in Gravitational Physics* (Cambridge Univ. Press, New York, ed. 2, 1993).
31. C. D. Hoyle *et al.*, *Phys. Rev. D* **70**, 042004 (2004).
32. See, for example, J. Polchinski, *String Theory* (Cambridge Univ. Press, New York, 1998), vol. 1.
33. N. Arkani-Hamed, S. Dimopoulos, G. Dvali, *Phys. Lett. B* **429**, 263 (1998).
34. L. Randall, R. Sundrum, *Phys. Rev. Lett.* **83**, 4690 (1999).
35. For a review, see L. Randall, *Science* **296**, 1422 (2002).
36. J. Garriga, T. Tanaka, *Phys. Rev. Lett.* **84**, 2778 (2000).
37. S. B. Giddings, E. Katz, L. Randall, *J. High Energy Phys.* **3**, 23 (2000).
38. P. Binetruy, C. Deffayet, D. Langlois, *Nucl. Phys. B* **565**, 269 (2000).
39. R. Maartens, D. Wands, B. A. Bassett, I. P. C. Heard, *Phys. Rev. D* **62**, 041301 (2000).
40. For a review, see D. Langlois, http://arxiv.org/abs/gr-qc/0410129.
41. F. Quevedo, *Class. Quant. Grav.* **19**, 5721 (2002).
42. A. D. Linde, http://arxiv.org/abs/hep-th/0402051.
43. C. P. Burgess, *Pramana* **63**, 1269 (2004).
44. G. R. Dvali, S.-H. H. Tye, *Phys. Lett. B* **450**, 72 (1999).
45. S. Kachru, R. Kallosh, A. Linde, S. P. Trivedi, *Phys. Rev. D* **68**, 046005 (2003).
46. S. Kachru *et al.*, *J. Cosmol. Astropart. Phys.* **0310**, 013 (2003).
47. N. Iizuka, S. P. Trivedi, *Phys. Rev. D* **70**, 043519 (2004).
48. R. Bousso, J. Polchinski, *J. High Energy Phys.* **0006**, 006 (2000).
49. L. Susskind, http://arxiv.org/abs/hep-th/0302219.
50. R. Bousso, J. Polchinski, *Sci. Am.* **291**, 60 (September 2004).
51. A. D. Linde, D. Linde, A. Mezhlumian, *Phys. Rev. D* **49**, 1783 (1994).
52. See, for example, J. Garriga, A. Vilenkin, *Phys. Rev. D* **64**, 023507 (2001) and also (*56*).
53. S. Weinberg, *Phys. Rev. Lett.* **59**, 2607 (1987).
54. T. Banks, M. Dine, E. Gorbatov, *J. High Energy Phys.* **0408**, 058 (2004).
55. For a review, see J. P. Ostriker, P. Steinhardt, *Science* **300**, 1909 (2003).
56. M. Tegmark, http://arxiv.org/abs/astro-ph/0410281.
57. We thank G. Dvali, H. Liu, M. Tegmark, S. Trivedi, and H. Tye for helpful comments on the manuscript. This work was supported in part by funds provided by the U.S. Department of Energy (D.O.E.) under cooperative research agreement no. DF-FC02-94ER40818.

10.1126/science.1107483

Genetic Consequences of Tropical Second-Growth Forest Regeneration

Uzay U. Sezen,* Robin L. Chazdon, Kent E. Holsinger

Despite a long history of international and local efforts, old-growth tropical forests are still being cleared. Today, second-growth forests cover more area than old-growth forests in many tropical countries (*1*). Yet we know little about the genetic composition of founding populations in second-growth forests (*2*). Parentage analysis can reveal patterns of gene flow through pollen and seed dispersal, providing accurate information about the sources of founding populations (*3*–*5*). Here, we assess the parentage of founders of the abundant canopy palm *Iriartea deltoidea* in a second-growth forest in Costa Rica.

Iriartea is an abundant canopy palm with a wide geographic distribution in neotropical forests (*6*). Its monoecious flowers are pollinated by stingless bees (*Trigona* spp. and *Melipona* spp.) and now by honeybees (*Apis* spp.). Seeds are dispersed by a variety of birds and mammals, including toucans, white-faced monkeys, peccaries, tapir, and several species of rodents. This shade-tolerant species shows broad representation of all size classes in mature forests.

We analyzed the entire founder generation of 130 flowering and fruiting trees in a 24-year-old second-growth forest adjacent to old-growth forest at La Selva Biological Field Station in Costa Rica (*7*). The study area was a 300 m by 1000 m protected area, covering 20 ha of second-growth and 10 ha of adjacent old-growth forest and protected since 1981. *Iriartea* depends on newly dispersed seed to recolonize second-growth forests. In our study area, all potential seed sources reside in old-growth forest. We defined the founder generation as the first generation derived from seeds that colonized abandoned pasture. Monitoring data obtained from a permanent sample plot show that *Iriartea* reached reproductive maturity in as little as 20 years in this second-growth forest. Our parentage analysis with 141 amplified fragment length polymorphism (AFLP) loci revealed high levels of reproductive dominance among parents of the source population (*7*). Among 66 old-growth forest trees, only two contributed 56% of the genes to the second-growth founder population; 23 contributed the remaining 44%; and 41 had no offspring.

Average seed dispersal distance greatly exceeded that of pollen dispersal. The median seed dispersal distance was 270 m, whereas the longest (minimum) dispersal distance we detected was more than 875 m (*7*). Maximum pollen movement was less than 220 m. The median pollination distance was 100 m, and 50% of the pollination events between 80 and 120 m corresponded to pollen movement between the two reproductively dominant trees.

Reproductive dominance of a few old-growth parents had strong genetic consequences in the second-growth forest. Genetic diversity in second-growth founders was significantly lower than in adjacent old-growth trees; mean dissimilarity among individuals was 0.58 in second-growth and 0.73 in old-growth forest (Tanimoto's *D*) (*7*) (Fig. 1). Patch diameters of similar genotypes were larger in second-growth forest (125 m) than in old-growth forest (96 m).

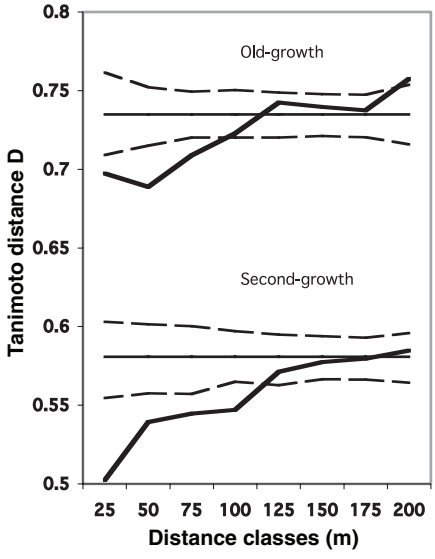


Fig. 1. Spatial autocorrelation distogram of mature trees in both forests using Tanimoto's *D* (bold lines). Confidence intervals (95%) for no spatial structure are indicated by dashed lines. Overall means of genetic distances for both forests are shown as thin flat lines as a reference for random spatial structure.

Reproductive dominance may be especially common in early successional forests. We do not know why two trees in our study show such extreme reproductive dominance. Even in this best-case scenario of forest regeneration, where numerous and diverse seed sources remain in a protected, adjacent, old-growth forest, where the animal disperser community is intact, and where there are no physical or ecological barriers to dispersal, more than half of the founding *Iriartea* population were full or half-siblings. Dominance patterns may be even more skewed in cases where the pool of parent trees is more restricted in size or genetic diversity.

Similar colonization patterns and genetic consequences may apply to a broad suite of tree and palm species that colonize young second-growth forest and persist as shade-tolerant species in later successional stages. Most studies of gene flow in plants have concentrated on pollen-mediated gene flow and male reproductive success (*3*, *4*). Colonization of second-growth forest depends on seed dispersal; here we demonstrate that dispersal can contribute more to total gene flow than pollen movement during tropical forest succession. We expect dispersal and (to a lesser extent) pollination distances to diminish as local seed production increases during succession.

Reduced genetic diversity of the founder population will likely persist in later generations as closely related founders contribute an increasing fraction of seedlings. Our results on the founders suggest that tree populations in second-growth forests will require continuous gene flow over successive generations to restore genetic diversity to levels currently observed in local old-growth forests.

References and Notes

1. Food and Agricultural Organization of the United Nations (FAO), *FAO For. Pap. 140* (FAO, Rome, 2001).
2. M. Cespedes, V. Gutierrez, N. M. Holbrook, O. J. Rocha, *Mol. Ecol.* **12**, 3201 (2003).
3. J. D. Nason, A. E. Herre, J. L. Hamrick, *Nature* **391**, 685 (1998).
4. M. R. Chase, C. Moller, R. Kesseli, K. S. Bawa, *Nature* **383**, 398 (1996).
5. P. R. Aldrich, J. L. Hamrick, *Science* **281**, 103 (1998).
6. A. Henderson, *Flora Neotrop. Monogr.* **53** (1990).
7. Materials and methods are available as supporting material on *Science* Online.
8. Supported by the Andrew W. Mellon Foundation.

Supporting Online Material

www.sciencemag.org/cgi/content/full/307/5711/891/DC1
Materials and Methods
References and Notes

9 September 2004; accepted 18 November 2004
10.1126/science.1105034

Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269–3043, USA.

*To whom correspondence should be addressed. E-mail: uzay.sezen@huskymail.uconn.edu

Twenty-One Millisecond Pulsars in Terzan 5 Using the Green Bank Telescope

Scott M. Ransom,[1,2]* Jason W. T. Hessels,[2] Ingrid H. Stairs,[3] Paulo C. C. Freire,[4] Fernando Camilo,[5] Victoria M. Kaspi,[2] David L. Kaplan[6]

We have identified 21 millisecond pulsars (MSPs) in globular cluster Terzan 5 by using the Green Bank Telescope, bringing the total of known MSPs in Terzan 5 to 24. These discoveries confirm fundamental predictions of globular cluster and binary system evolution. Thirteen of the new MSPs are in binaries, of which two show eclipses and two have highly eccentric orbits. The relativistic periastron advance for the two eccentric systems indicates that at least one of these pulsars has a mass 1.68 times greater than the mass of the Sun at 95% confidence. Such large neutron star masses constrain the equation of state of matter at or beyond the nuclear equilibrium density.

The extremely high stellar densities (10^4 to 10^6 pc^{-3}) in the cores of globular clusters (GCs) result in stellar interactions that produce and destroy binary systems as well as exchange their members (*1*). Formed by the death of massive stars early in a cluster's history, neutron stars (NSs) usually reside near the cores of clusters because of their relatively large masses and the mass segregation induced by dynamical friction. There they are likely to interact with one or more stars over the 10^{10}-year lifetimes of GCs. These interactions lead to a production rate of low-mass x-ray binaries (LMXBs) and their progeny such as MSPs [via the recycling mechanism (*2*)] that is highly enhanced compared with the rate in the Milky Way Galaxy. Before our observations, there were 80 pulsars (most of them binary MSPs) known in 24 GCs (*3*), with the relatively massive and nearby cluster 47 Tucanae containing 22 of these (*4*). Finding and monitoring many pulsars in a single cluster provides unique probes into a range of GC, binary evolutionary, and stellar astrophysics (*5*, *6*). In addition, GCs produce exotic systems such as highly eccentric (*7*) and MSP−main sequence star binaries (*8*).

The interaction rate between NSs and other stars or binaries in a GC is a complex function of the total cluster mass, the size of the cluster core (*9*) and its stellar density, the initial stellar mass function, and the level of mass segregation present in the cluster (*10*). However, relatively simple theoretical modeling of stellar interaction rates (*11*), as well as one known LMXB and several additional x-ray sources detected recently with the Chandra X-ray Observatory (*12*), indicate that the dense, massive, and metal-rich GC Terzan 5 has one of the highest stellar interaction rates of any cluster in the Galaxy (*13*) and perhaps also the largest number of MSPs (*14*, *15*). But because Terzan 5 [Galactic coordinates (l, b) = (3.8°, 1.7°)] is distant (D = 8.7 ± 2 kpc) and located within ~1 kpc of the Galactic center (*16*), the large column density of interstellar free electrons (i.e., the dispersion measure, DM ~ 240 pc cm^{-3}) produces considerable dispersive smearing ($\propto \nu^{-3}$ for typical pulsar search data) and scatter broadening ($\propto \nu^{-4.4}$) of radio pulses, which hinders pulsation searches for MSPs at the often-observed radio frequencies (ν) of 400 to 1400 MHz.

Deep radio images of Terzan 5, which are not affected by the dispersive effects of the interstellar medium (ISM), were made with the Very Large Array (VLA) and showed what appeared to be the integrated emission from many tens or even hundreds of pulsars with low flux densities (*17*). Surprisingly, numerous deep searches for pulsars over the past 15 years using the Parkes radio telescope at 1400 MHz have identified only three (*18*): the 11-ms Ter 5 A with its unusual eclipses (*19*) and the two isolated MSPs Ter 5 C (*20*) and D (*21*).

Observations and data analysis. On 17 July 2004, we observed Terzan 5 for 5.9 hours with the National Radio Astronomy Observatory's (*22*) 100-m Green Bank Telescope (GBT) and the Pulsar Spigot backend (*23*). The S-band receiver provided 600 MHz (1650 to 2250 MHz) of relatively interference-free bandwidth in two orthogonal polarizations, which the Spigot summed and synthesized into 768 0.78125-MHz frequency channels every 81.92 μs. With the known DM and scattering time scale (*24*) toward the cluster, the effective time resolution of the data was ~0.3 ms. We observed only a single position, because the 6.5′ telescope beam width at these frequencies is much larger than the 0.83′ half-mass radius of the cluster (*25*). The large unblocked aperture of the GBT, the wide bandwidth provided by the Spigot and the S-band receiver, and the move to higher observing frequencies together provided an increase in sensitivity over the Parkes searches (*21*) by factors of ~5 for typical recycled pulsars and by ≳10 for MSPs with spin periods $P_{psr} \lesssim 2$ ms (*26*).

We searched the observation by dedispersing the raw data into 40 separate time series with DMs ranging from 230 to 250 pc cm^{-3} and spaced by 0.5 pc cm^{-3}. We Fourier-transformed the full 5.9-hour time series as well as 10-, 20-, and 60-min sections and searched them by using Fourier-domain acceleration search techniques (*27*) in a manner similar to that described in (*28*) in order to maintain sensitivity to pulsars in compact binary systems. In that first observation, we detected 14 new pulsars, Ter 5 E to R (*29*). By using eight more observations taken between July and November 2004, we found seven additional pulsars and determined the basic orbital parameters of 10 of the 13 new binaries (Fig. 1 and Table 1). In general, the uncalibrated flux densities of the pulsars (as computed by comparing the integrated signal from a pulsar to the predicted total system noise level) were constant to within a level of ≲50% between these observations, implying that diffractive scintillation will not largely affect pulsar measurements for Terzan 5 at these observing frequencies. Several additional binary pulsar candidates from these data remain unconfirmed but may reappear in future observations at more fortuitous orbital phases. Because the volume enclosed by the GBT beam at 1950 MHz out to a distance of 10 kpc is ~10^{-3} kpc^3 and it has been estimated (*30*) that there are roughly 10^5 observable MSPs within the Galaxy (with volume ~ 10^2 to 10^3 kpc^3), it is possible that a detectable foreground MSP is within our beam. However, given that the DM toward Terzan 5 is known

[1]National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903, USA. [2]Department of Physics, McGill University, 3600 University Street, Montreal, QC H3A 2T8, Canada. [3]Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, BC V6T 1Z1, Canada. [4]National Astronomy and Ionosphere Center, Arecibo Observatory, HC03 Box 53995, PR 00612, USA. [5]Columbia Astrophysics Laboratory, Columbia University, 550 West 120th Street, New York, NY 10027, USA. [6]Center for Space Research, Massachusetts Institute of Technology, 70 Vassar Street, Cambridge, MA 02139, USA.

*To whom correspondence should be addressed. E-mail: sransom@nrao.edu

to within ~5%, the probability of that MSP having a similar DM is much smaller, and hence we are confident that all of the new pulsars are members of Terzan 5. Positions with arc sec accuracy of the new MSPs from pulsar timing will solidify the associations.

The new pulsar population. With our initial discovery data, the ensemble of pulsars all have measured DMs with errors $\lesssim$ 0.1 pc cm^{-3}, flux densities, S_{1950}, with fractional errors of ~30%, and precisely determined P_{psr} values. These measurements allow us to compare the new systems with known MSPs such as those in 47 Tuc. Once precise positions and pulsar spin period derivatives have been determined from timing observations over the next year (the positions of the new pulsars are currently known only to within the 6.5′ primary beam of the GBT), a variety of tests of the gravitational potential and internal dynamics of Terzan 5 will be possible, in addition to estimates of the ages and magnetic field strengths of many of the pulsars (*31*).

The DMs for the pulsars in Terzan 5 are distributed with a central Gaussian-like component having an average and standard deviation of 237.8 ± 1.4 pc cm^{-3} and five outliers having higher and lower DMs (Table 1). The overall spread in DM, 9.5 pc cm^{-3} (13 times larger than what is observed for 47 Tuc), is the largest known for any GC and is likely due to the large average DM for Terzan 5 and irregularities in the ISM along the differing sight lines toward its pulsars. The Gaussian-like component of the DM distribution likely corresponds to a group of centrally concentrated pulsars near the cluster core. The position of Ter 5 C, which has an average DM, is about 10″ or one core radius (*9*) north of the cluster center and supports such a notion. Likewise, pulsars with outlying DMs, such as Ter 5 D and J, are probably more offset like Ter 5 A, which has a high DM and is located 36″ from the center (*20*).

The VLA radio imaging of Terzan 5 (*17*) revealed numerous point sources within 30″ of the cluster center, as well as ~2 mJy (1 Jy = 10^{-26} W m^{-2} Hz^{-1}) of diffuse emission at 1400 MHz within ~10″. The central emission was attributed to 60 to 200 unresolved pulsars assuming D = 7.1 kpc, a standard luminosity distribution, and a minimum pulsar luminosity ($L_{1400} \equiv S_{1400}D^2$) of $L_{1400,min}$ = 0.3 mJy kpc^2. A distance to Terzan 5 of D = 8.7 kpc (*16*) would increase the number of unresolved pulsars by ~50%. For a typical pulsar radio spectral index of −1.6 (*32*) and omitting the flux densities of Ter 5 A and C, which are located outside the region of diffuse emission (*20*), the integrated measured flux density from the other 22 pulsars at 1400 MHz is ~1.3 mJy. If the five pulsars with outlying DMs reside outside the cluster core, the total flux density from the remaining pulsars is only ~1 mJy. None of the new pulsars can individually (or in combination with nearby Ter 5 C) account for the bright (1.42 mJy) point source ''N'' located 12″ north of the cluster center (*17*). However, given its very wide pulse profile, lack of a flat off-pulse baseline, and positional coincidence (*20*), Ter 5 C could possibly emit substantial unpulsed emission like MSP J0218+4232 (*33*) and thereby account for all of the flux density of source N.

The differential luminosity distribution of the new pulsars resembles that of the 47 Tuc (*4*), M15 (*5*), and the Galactic disk (*34*) populations by following the normal $d\log N =$

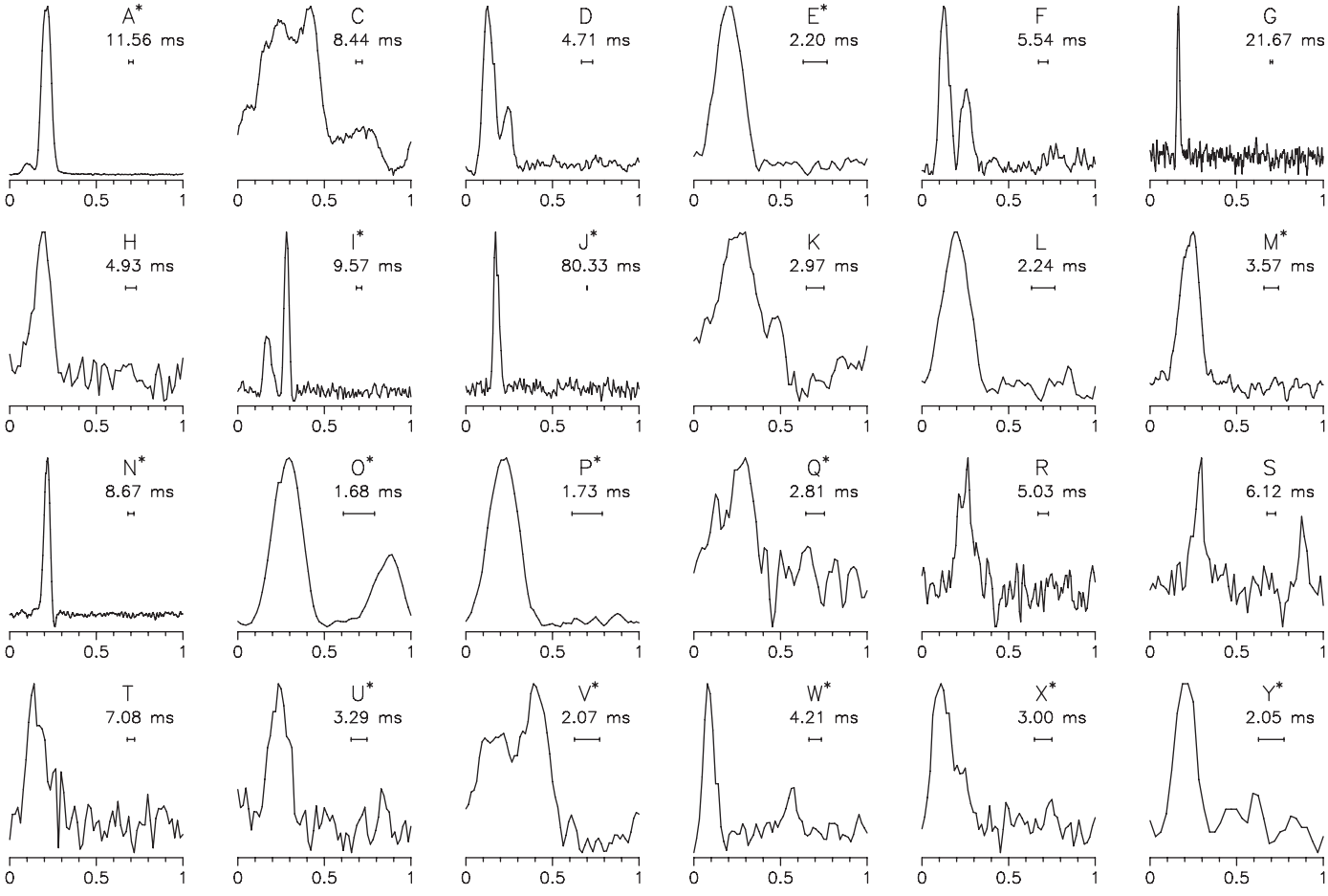


Fig. 1. The 1950 MHz GBT+Spigot pulse profiles for each of the pulsars known in Terzan 5. All but Ter 5 A, C, and D are newly discovered (*18*). Each profile is the weighted average of the best detections of that pulsar and is a measure of the relative flux density as a function of rotational phase. Asterisks indicate that the pulsar is the member of a binary system, and the length of the horizontal error bar (0.3 ms) is the effective system time resolution.

$-d\log L$ relation, with no sign of a downturn at the low luminosity end ($L_{1400} \sim 1$ to 2 mJy kpc^2). This implies that we have not reached the lowest luminosities of the intrinsic pulsar distribution in Terzan 5, nor have we yet reached the sensitivity limit of the observing system (*35*). It is likely that tens of less-luminous pulsars remain to be discovered in Terzan 5, possibly even several bright ones (e.g., the source N). Such bright pulsars may remain undetected if they are members of very compact or massive binaries due to eclipsing or excessive doppler accelerations or if they have extremely fast spin periods ($P_{psr} \ll 1.5$ ms). Detailed searches using more advanced search techniques may yet uncover these systems.

The 24 known pulsars in Terzan 5 and the 22 in 47 Tuc appear to have different spin period distributions. The 47 Tuc pulsars are a homogeneous population with periods from 2.1 to 7.6 ms (*3*), whereas those in Terzan 5 have a flatter distribution that includes six pulsars slower than 7.6 ms and the four fastest pulsars known in GCs (P_{psr} values of 1.67, 1.73, 2.05, and 2.07 ms). A Kolmogorov-Smirnov test suggests that these samples are drawn from different parent distributions at 85% confidence. These differences may be related to the dynamical states of the cluster cores [i.e., just pre- or just postcore collapse for Terzan 5 (*16*)], different epochs of MSP creation, or the occurrence of unusual evolutionary mechanisms that only manifest themselves at very high (i.e., $>10^5$ pc^{-3}) stellar densities.

Individual pulsars. The binaries Ter 5 E and W have orbital periods, P_{orb}, of ~60 days and ~4.9 days, respectively. In the dense stellar environment of a GC, such wide binaries have large cross sections for stellar encounters that can disrupt them, eject them from the cluster core, or, after multiple collisions, induce eccentricity (*36*) in their initially circular orbits (*31*). For these pulsars, the measured eccentricities are $e \sim 0.02$, substantially larger than those predicted for binary MSPs with helium white dwarf (WD) companions in the Galaxy (*31*). If they are located near the cluster center, where the stellar densities are $\gtrsim 10^5$ pc^{-3} (*25*), the time scale for interactions to produce such eccentricities is 10^8 to 10^9 years (*36*), consistent with the $\gtrsim 10^9$-year ages of most cluster pulsars. However, the 60-day orbit of Ter 5 E implies a time scale near the low end of that range, possibly too short for MSP lifetimes. This may indicate that it resides further from the center of Terzan 5: More than 10^9 years spent in the core could result in many interactions that would either destroy the binary or induce much larger eccentricities. The eccentricities of Ter 5 I and J are too large to have been produced by this method, especially given the relatively compact nature of their orbits, indicating that they were formed by a different mechanism.

Table 1. Known pulsars in Terzan 5 (*18*). Pulsars listed without orbital parameters are likely isolated systems, whereas those marked with an *e* are eclipsing systems. The errors on the dispersion measures (DMs) range from 0.01 to 0.1 pc cm^{-3}, and the errors on the measured flux densities are ~30%. The flux densities for the eclipsing pulsars include only the times when the pulsar is not eclipsed. The light travel time across the projected pulsar semimajor axis is defined as $x \equiv a_1 \sin(i)/c$. Eccentricities listed as "0" are too small to measure at present and have been set to zero for orbital parameter fitting. The minimum companion mass m_2 was calculated assuming a pulsar mass m_1 of 1.4 $M_\odot$ and $i = 90°$ except for Ter 5 I and J (Fig. 2). All measured parameters were determined with the use of the Tempo software package (*47*). Unk., unknown.

PSR	P_{psr} (ms)	Dispersion measure (pc cm^{-3})	1950 MHz flux density (μJy)	P_{orb} (days)	x (lt-s)	Eccentricity	Minimum m_2 ($M_\odot$)
A[e]	11.56315	242.44	1020	0.0756	0.120	0	0.089
C	8.43610	237.14	360				
D	4.71398	243.83	41				
E	2.19780	236.84	48	60.06	23.6	~ 0.02	0.22
F	5.54014	239.18	35				
G	21.67187	237.57	15				
H	4.92589	238.13	15				
I	9.57019	238.73	29	1.328	1.818	0.428	0.24
J	80.33793	234.35	19	1.102	2.454	0.350	0.38
K	2.96965	234.81	40				
L	2.24470	237.74	41				
M	3.56957	238.65	33	0.4431	0.596	0	0.14
N	8.66690	238.47	55	0.3855	1.619	0.000045	0.48
O[e]	1.67663	236.38	120	0.2595	0.112	0	0.036
P[e]	1.72862	238.79	77	0.3626	1.272	0	0.38
Q	2.812	234.50	27	> 1?	Unk.	Unk.	Unk.
R	5.02854	237.60	12				
S	6.11664	236.26	18				
T	7.08491	237.70	20				
U	3.289	235.50	16	> 1?	Unk.	Unk.	Unk.
V	2.07251	239.11	71	0.5036	0.567	0	0.12
W	4.20518	239.14	22	4.877	5.869	0.015	0.30
X	2.999	240.03	18	> 1?	Unk.	Unk.	Unk.
Y	2.04816	239.11	16	1.17	1.16	0	0.14

At least two of the new binary pulsars, Ter 5 O and P, have been observed to eclipse, although given the factor of 10 difference in the inferred companion masses ($m_2 \gtrsim 0.036\ M_\odot$ for Ter 5 O and $\gtrsim 0.38\ M_\odot$ for Ter 5 P, where $M_\odot$ is the mass of the Sun) and eclipse durations ($\sim 0.05 P_{orb}$ for Ter 5 O and $\gtrsim 0.5 P_{orb}$ for Ter 5 P), the systems are unalike. Ter 5 O, the third fastest MSP known, is similar to the Black-Widow eclipsing MSP B1957+20 (*37*), with a very low-mass companion and an eclipse duration (~ 15 to 20 min) corresponding to a physical size (~0.03 $R_\odot$, where $R_\odot$ is the radius of the Sun) much larger than the companion star's Roche lobe. Such systems are common in GCs (*38*). Ter 5 P, the fourth fastest MSP known, is unusual and most similar to the binary MSP NGC 6397 A, whose peculiar red straggler or sub-subgiant companion causes irregular and long-duration eclipses of the radio pulses (*8*). Ter 5 P's eclipses are also irregular and correspond to an eclipsing region of several solar radii in size, indicating the companion is likely a peculiar evolved star as well. If so, the system may have been created after at least one exchange encounter where the MSP's original companion (which had spun up the pulsar) was ejected and replaced by a main sequence star. Such an encounter might eject the system from the core of the cluster, as has been observed for NGC 6397 A [although alternative formation scenarios for this pulsar have been proposed as well (*39*)].

The pulsars Ter 5 I and J are in highly eccentric orbits with companions of at least 0.24 and 0.38 $M_\odot$, respectively (Fig. 2). The companions are not sufficiently massive to be NSs unless the systems have improbably low inclination angles i (i.e., are almost face-on). If they were main-sequence or giant stars, eclipses would likely be observed for a variety of inclinations given the compactness of the orbits (orbital separations of ~5 $R_\odot$). Additionally, the orbital circularization time scales would be on the order of 10^5 years for both pulsars and so no eccentricity should currently be observed (*40*). The companions are therefore probably WDs. However, the pulsar recycling scenario generates MSP-WD binaries in nearly circular orbits because of tidal interactions during mass transfer and therefore did not produce these eccentric systems. A possible scenario involves the off-center collision of a NS with a red giant, which would disrupt the giant's envelope

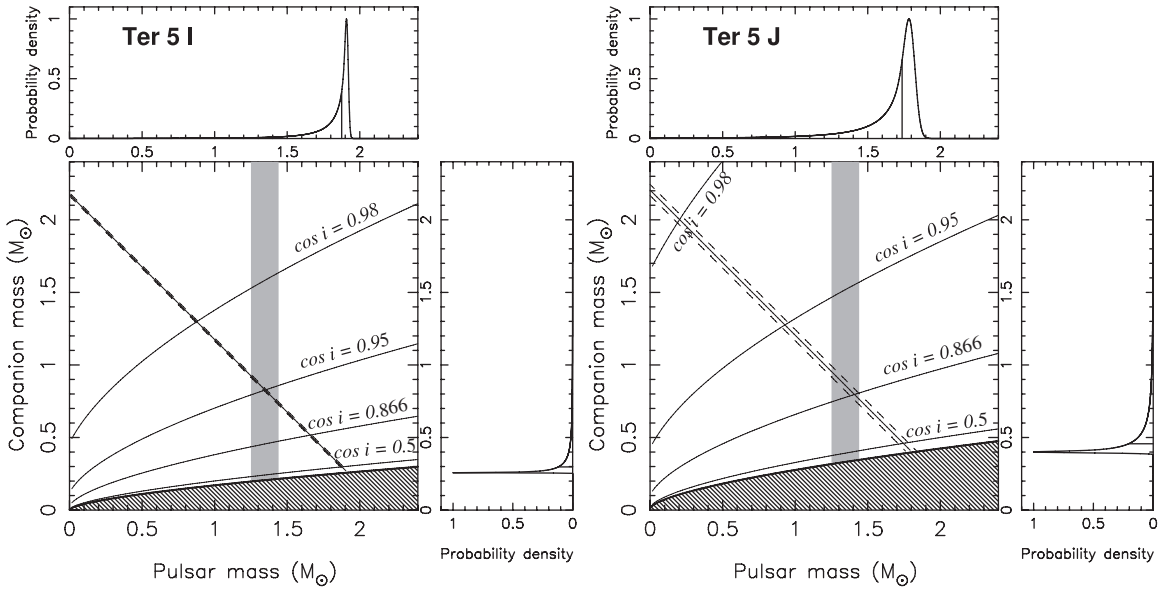


Fig. 2. Pulsar mass versus companion mass diagrams for the two highly eccentric binary pulsars Ter 5 I (**left**) and Ter 5 J (**right**). The hatched regions are excluded because of the definition of the Keplerian mass function (i.e., $\sin i \leq 1$). The diagonal band in the center of the figure shows the total system mass with 1σ confidence intervals as measured by the general-relativistic advance of periastron $\dot{\omega} = 0.255 \pm 0.001°$ year^{-1} for I and $\dot{\omega} = 0.327 \pm 0.004°$ year^{-1} for J. The marginal distributions show the masses of the pulsars and companions assuming a random distribution of inclinations (i.e., with probability density flat in cos *i*). The solid curves in the main plot indicate inclinations of (from top to bottom) 11.4°, 18.2°, 30°, 60°, and 90°. The gray vertical band shows the range of precisely measured NS masses from relativistic binary radio pulsars (*46*). In both cases, the median pulsar mass (indicated by the vertical line in the marginal distribution) lies significantly above 1.7 $M_\odot$, implying that one or both of these pulsars is considerably more massive than the NSs that have been well measured to date. A strict upper limit to the masses of both pulsars is 1.96 $M_\odot$.

and leave the core (now the WD) in an eccentric orbit about the NS (*41*). An alternative scenario involves the exchange of an initially isolated WD into a binary consisting of an MSP and the low-mass WD that recycled it. However, the observed pulsars are only mildly recycled. Furthermore, the multiple exchanges required in this scenario imply that the system was created near the core, where because of mass segregation the isolated WDs would be more massive than the observed companions to Ter 5 I and J. We therefore regard the collision scenario as more likely.

The high eccentricities of both orbits permitted us to measure their advances of the angle of periastron, $\dot{\omega} = 0.255 \pm 0.001°$ year^{-1} for I and $0.327 \pm 0.004°$ year^{-1} for J, which are likely caused by general-relativistic modification of the Keplerian elliptical orbit. Under this assumption, the total mass can be derived for each system: $2.17 \pm 0.02\ M_\odot$ for I and $2.20 \pm 0.04\ M_\odot$ for J. Together with the measured Keplerian mass functions in both cases, this result implies that the most likely NS masses are $> 1.7\ M_\odot$ (Fig. 2). However, classical contributions to $\dot{\omega}$ from tidal or rotational quadrupole mass moments in the companion star must also be considered (*42*, *43*). Tidal deformation is insignificant for a WD companion (*42*), but rotationally induced quadrupoles are possible if the WD is rapidly rotating, because there is no reason to expect the spin axis of the companion to be aligned with the orbital angular momentum.

The predicted contributions to $\dot{\omega}$ for Ter 5 I and J due to a rotationally induced quadrupole are $\dot{\omega}_{\rm rot} \sim 0.01$ to $0.02°$ year^{-1} times angular and stellar structure factors (*42*–*44*). One of the stellar structure factors, α_6, can range up to 15 for WDs (*42*), making $\dot{\omega}_{\rm rot}$ a potentially considerable contribution to the measured $\dot{\omega}$. A substantial $\dot{\omega}_{\rm rot}$, however, will produce changes in *i* and hence in the projected semimajor axis $x \equiv a_1 \sin(i)/c$. The size of this effect may be written as $\dot{\omega}_{\rm rot} \sim \dot{x}_{\rm rot}/x$ times a trigonometric factor usually of order unity that depends on *i*, the angle between the WD rotation axis and the orbital angular momentum vector, and the phase of the orbital precession (*43*). For random choices of these angles, the trigonometric factor is <10 ~80% of the time. By incorporating detections from Parkes search observations taken in 1998 and 2000 (*21*) with the use of the now-known orbital ephemerides, we have upper limits on $\dot{x}$, which imply that $\dot{\omega}_{\rm rot}$ is $\lesssim 0.003°$ year^{-1} times the trigonometric factor for both systems. The magnitudes of these rotational-quadrupole contributions are comparable to our current measurement uncertainties for $\dot{\omega}$. We emphasize that this result does not depend on the type of companion and is equally valid for main-sequence stars and WDs. Therefore, unless the orientations of the orbital and rotational angular momentum vectors are fine-tuned for both systems, the $\dot{\omega}$ measurements are well described by general relativity. Under this assumption, one or both of these NSs are likely to be unprecedentedly massive: calculation of the joint probabilities (Fig. 2) indicates that at least one of the pulsars is more massive than 1.48, 1.68, or 1.74 $M_\odot$ at 99%, 95%, and 90% confidence levels, respectively. A strict upper limit to the masses of both pulsars is 1.96 $M_\odot$. Similar but slightly less stringent evidence from pulsar timing for a massive NS in the Galactic low-mass WD-MSP binary J0751+1807 has been presented (*45*). It is an intriguing question whether the large masses for Ter 5 I and J resulted from the collision formation mechanism, possibly during a short period of hypercritical accretion which partially or completely spun-up the pulsars (*41*). GBT observations over the next 1 to 3 years will measure the varying delays due to gravitational redshift and time dilation as Ter 5 I moves in its orbit, which together are known as γ (the measurement of γ for Ter 5 J will take 5 to 10 years because of the relatively slow spin period of the pulsar). This parameter, along with general relativity and the already well-measured $\dot{\omega}$, will provide a precise mass for the pulsar and will likely rule out several soft equations of state for matter at nuclear densities (*46*).

References and Notes

1. G. Meylan, D. C. Heggie, *Astron. Astrophys. Rev.* **8**, 1 (1997).
2. M. A. Alpar, A. F. Cheng, M. A. Ruderman, J. Shaham, *Nature* **300**, 728 (1982).
3. For an updated list, see www.naic.edu/~pfreire/GCpsr.html.
4. F. Camilo, D. R. Lorimer, P. Freire, A. G. Lyne, R. N. Manchester, *Astrophys. J.* **535**, 975 (2000).
5. S. B. Anderson, thesis, California Institute of Technology, Pasadena, CA (1992).
6. P. C. Freire *et al.*, *Mon. Not. R. Astron. Soc.* **340**, 1359 (2003).
7. P. C. Freire, Y. Gupta, S. M. Ransom, C. H. Ishwara-Chandra, *Astrophys. J.* **606**, L53 (2004).
8. N. D'Amico *et al.*, *Astrophys. J.* **561**, L89 (2001).
9. The core radius is defined as that where the optical surface brightness is half its central value.
10. F. Verbunt, in *New Horizons in Globular Cluster Astronomy*, F. Piotto, G. Meylan, S. G. Djorgovski, M. Riello, Eds. [Astronomical Society of the Pacific (ASP) Conference Series no. 296, ASP, San Francisco, CA, 2003], p. 245.

11. F. Verbunt, in *Omega Centauri, A Unique Window into Astrophysics*, F. van Leeuwen, J. D. Hughes, F. Piotto, Eds. (ASP Conference Series no. 265, ASP, San Francisco, CA, 2002), p. 289.
12. C. O. Heinke *et al.*, *Astrophys. J.* **590**, 809 (2003).
13. D. Pooley *et al.*, *Astrophys. J.* **591**, L131 (2003).
14. S. Sigurdsson, E. S. Phinney, *Astrophys. J. Suppl. Ser.* **99**, 609 (1995).
15. S. R. Kulkarni, S. B. Anderson, in *Dynamical Evolution of Star Clusters–Confrontation of Theory and Observations*, P. Hut, J. Makino, Eds. (International Astronomical Union Symposium no. 174, Kluwer, Dordrecht, Netherlands, 1996), p. 181.
16. H. N. Cohn, P. M. Lugger, J. E. Grindlay, P. D. Edmonds, *Astrophys. J.* **571**, 818 (2002).
17. A. S. Fruchter, W. M. Goss, *Astrophys. J.* **536**, 865 (2000).
18. The 442-ms pulsar J1748−2444 was initially identified as Ter 5 B (B1744−24B) but is now known to be a foreground pulsar unrelated to Terzan 5 (*20*).
19. A. G. Lyne *et al.*, *Nature* **347**, 650 (1990).
20. A. G. Lyne, S. H. Mankelow, J. F. Bell, R. N. Manchester, *Mon. Not. R. Astron. Soc.* **316**, 491 (2000).
21. S. M. Ransom, thesis, Harvard University, Cambridge, MA (2001).
22. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Incorporated.
23. D. Kaplan *et al.*, in preparation.
24. D. J. Nice, S. E. Thorsett, *Astrophys. J.* **397**, 249 (1992).
25. W. E. Harris, *Astron. J.* **112**, 1487 (1996).
26. These comparisons assume a typical pulsar spectral index of −1.6 (*32*).
27. S. M. Ransom, S. S. Eikenberry, J. Middleditch, *Astron. J.* **124**, 1788 (2002).
28. S. M. Ransom *et al.*, *Astrophys. J.* **604**, 328 (2004).
29. Ter 5 E was a candidate (i.e., unconfirmed) pulsar in (*21*).
30. A. G. Lyne *et al.*, *Mon. Not. R. Astron. Soc.* **295**, 743 (1998).
31. E. S. Phinney, S. R. Kulkarni, *Annu. Rev. Astron. Astrophys.* **32**, 591 (1994).
32. D. R. Lorimer, J. A. Yates, A. G. Lyne, D. M. Gould, *Mon. Not. R. Astron. Soc.* **273**, 411 (1995).
33. J. Navarro, G. de Bruyn, D. Frail, S. R. Kulkarni, A. G. Lyne, *Astrophys. J.* **455**, L55 (1995).
34. A. G. Lyne, R. N. Manchester, J. H. Taylor, *Mon. Not. R. Astron. Soc.* **213**, 613 (1985).
35. $S_{1950,\mathrm{min}}$ is about 8 μJy for $P \sim$ 2- to 4-ms isolated or long orbital period MSPs.
36. F. A. Rasio, D. C. Heggie, *Astrophys. J.* **445**, L133 (1995).
37. A. S. Fruchter, D. R. Stinebring, J. H. Taylor, *Nature* **333**, 237 (1988).
38. A. R. King, M. B. Davies, M. E. Beer, *Mon. Not. R. Astron. Soc.* **345**, 678 (2003).
39. A. Possenti *et al.*, in *Binary Radio Pulsars*, F. Rasio, I. Stairs, Eds. (ASP Conference Series no. 328, ASP, San Francisco, CA, 2005), p. 189.
40. J. Tassoul, *Astrophys. J.* **444**, 338 (1995).
41. F. A. Rasio, S. A. Shapiro, *Astrophys. J.* **377**, 559 (1991).
42. L. L. Smarr, R. Blandford, *Astrophys. J.* **207**, 574 (1976).
43. N. Wex, *Mon. Not. R. Astron. Soc.* **298**, 997 (1998).
44. E. M. Splaver *et al.*, *Astrophys. J.* **581**, 509 (2002).
45. D. J. Nice, E. M. Splaver, I. H. Stairs, in *Binary Radio Pulsars*, F. Rasio, I. Stairs, Eds. (ASP Conference Series no. 328, ASP, San Francisco, CA, 2005), p. 371.
46. J. M. Lattimer, M. Prakash, *Science* **304**, 536 (2004).
47. The Tempo program is available online at http://pulsar.princeton.edu/tempo.

48. We thank F. Rasio, S. Sigurdsson, and M. van Kerkwijk for extremely useful discussions and J. Herrnstein, L. Greenhill, D. Manchester, A. Lyne, and N. D'Amico for providing or aiding with the Parkes data from 1998 and 2000. J.W.T.H. is a Natural Sciences and Engineering Research Council of Canada (NSERC) Post-Graduate Scholarship–Doctoral fellow. I.H.S. holds an NSERC University Faculty Award and is supported by a Discovery grant and University of British Columbia start-up funds. F.C. thanks support from NSF. V.M.K. holds a Canada Research Chair and is supported by an NSERC Discovery Grant and Steacie Fellowship Supplement, by the Fonds Québécois de la recherche sur la nature et les technologies and Canadian Institute for Advanced Research, and by a New Opportunities Grant from the Canada Foundation for Innovation. D.L.K. is a Pappalardo Fellow.

13 December 2004; accepted 4 January 2005
Published online 13 January 2005;
10.1126/science.1108632
Include this information when citing this paper.

Optical Imaging of Neuronal Populations During Decision-Making

K. L. Briggman,[1] H. D. I. Abarbanel,[2,3] W. B. Kristan Jr.[1]*

We investigated decision-making in the leech nervous system by stimulating identical sensory inputs that sometimes elicit crawling and other times swimming. Neuronal populations were monitored with voltage-sensitive dyes after each stimulus. By quantifying the discrimination time of each neuron, we found single neurons that discriminate before the two behaviors are evident. We used principal component analysis and linear discriminant analysis to find populations of neurons that discriminated earlier than any single neuron. The analysis highlighted the neuron cell 208. Hyperpolarizing cell 208 during a stimulus biases the leech to swim; depolarizing it biases the leech to crawl or to delay swimming.

Understanding the mechanisms of behavioral choice would be a major step in bringing together neuroscience, psychology, and ethology (*1*). Research into decision-making has used several different strategies. One very productive approach is to have a primate make a sensory discrimination between very similar stimuli while the activity of neurons in various parts of the nervous system is recorded (*2*–*8*). A second approach uses choice competition: presenting an animal with two stimuli that produce mutually exclusive behaviors (choices), to see which behavior predominates (*9*). This has led to the notion that behavioral choices are hierarchical. The neuronal mechanism originally proposed to underlie behavioral hierarchies was inhibitory interactions among the neurons responsible for triggering the different behaviors (*10*). Later work has found that neurons capable of eliciting one behavior are often activated during other, sometimes conflicting, behaviors (*11*, *12*). Among other things, this observation suggests that individual decision-making neurons can be multiplexed—they contribute to choosing more than one behavior—and that they trigger behaviors by being active with other combinations of neurons.

We used a third approach to study decision-making: choice variability. We presented a nervous system with identical stimuli that repeatedly produce two different, mutually exclusive behaviors with roughly equal probabilities. This approach allowed us to focus on neurons involved in decision-making that are downstream from neurons used to make sensory discriminations.

We used the isolated central nervous system (CNS) of the medicinal leech. Motor neuron activity patterns characteristic of swimming (*13*) and crawling (*14*) can be elicited from isolated preparations by electrically stimulating peripheral nerves. Such sensory stimulation activates mechanosensory neurons in patterns that mimic touching the leech's skin (*15*). Stimulating the same kinds of mechanosensory neurons in different locations on the leech produces characteristic behaviors like swimming or crawling (*16*, *17*). We follow the terminology proposed by Schall (*18*), referring to the different behavioral outputs as choices and the process leading up to a choice as decision-making.

Previously, recording from neurons intracellularly one at a time, then stimulating them to determine their effect on the initiation of behavior, has successfully uncovered interneurons that activate swimming (*19*, *20*), crawling (*21*), and whole-body shortening (*22*). However, to explore how decisions are made by populations of neurons (*11*), we needed to record from many neurons at once (*23*). We therefore used voltage-sensitive dyes (*24*, *25*) that allowed us to record simultaneously from many neurons in a midbody segmental ganglion at a resolution better than 5 mV (*26*).

The leech CNS makes behavioral choices. The isolated leech CNS consists of a nerve cord connecting 21 segmental ganglia plus a head and tail brain (Fig. 1A). This preparation generates motor patterns that are recognizable as behaviors observed in in-

[1]Division of Biological Sciences, [2]Department of Physics, University of California–San Diego, La Jolla, CA 92093–0357, USA. [3]Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, CA 92093–0402, USA.

*To whom correspondence should be addressed. E-mail: wkristan@ucsd.edu

11. F. Verbunt, in *Omega Centauri, A Unique Window into Astrophysics*, F. van Leeuwen, J. D. Hughes, F. Piotto, Eds. (ASP Conference Series no. 265, ASP, San Francisco, CA, 2002), p. 289.
12. C. O. Heinke *et al.*, *Astrophys. J.* **590**, 809 (2003).
13. D. Pooley *et al.*, *Astrophys. J.* **591**, L131 (2003).
14. S. Sigurdsson, E. S. Phinney, *Astrophys. J. Suppl. Ser.* **99**, 609 (1995).
15. S. R. Kulkarni, S. B. Anderson, in *Dynamical Evolution of Star Clusters–Confrontation of Theory and Observations*, P. Hut, J. Makino, Eds. (International Astronomical Union Symposium no. 174, Kluwer, Dordrecht, Netherlands, 1996), p. 181.
16. H. N. Cohn, P. M. Lugger, J. E. Grindlay, P. D. Edmonds, *Astrophys. J.* **571**, 818 (2002).
17. A. S. Fruchter, W. M. Goss, *Astrophys. J.* **536**, 865 (2000).
18. The 442-ms pulsar J1748−2444 was initially identified as Ter 5 B (B1744−24B) but is now known to be a foreground pulsar unrelated to Terzan 5 (*20*).
19. A. G. Lyne *et al.*, *Nature* **347**, 650 (1990).
20. A. G. Lyne, S. H. Mankelow, J. F. Bell, R. N. Manchester, *Mon. Not. R. Astron. Soc.* **316**, 491 (2000).
21. S. M. Ransom, thesis, Harvard University, Cambridge, MA (2001).
22. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Incorporated.
23. D. Kaplan *et al.*, in preparation.
24. D. J. Nice, S. E. Thorsett, *Astrophys. J.* **397**, 249 (1992).
25. W. E. Harris, *Astron. J.* **112**, 1487 (1996).
26. These comparisons assume a typical pulsar spectral index of −1.6 (*32*).
27. S. M. Ransom, S. S. Eikenberry, J. Middleditch, *Astron. J.* **124**, 1788 (2002).
28. S. M. Ransom *et al.*, *Astrophys. J.* **604**, 328 (2004).
29. Ter 5 E was a candidate (i.e., unconfirmed) pulsar in (*21*).
30. A. G. Lyne *et al.*, *Mon. Not. R. Astron. Soc.* **295**, 743 (1998).
31. E. S. Phinney, S. R. Kulkarni, *Annu. Rev. Astron. Astrophys.* **32**, 591 (1994).
32. D. R. Lorimer, J. A. Yates, A. G. Lyne, D. M. Gould, *Mon. Not. R. Astron. Soc.* **273**, 411 (1995).
33. J. Navarro, G. de Bruyn, D. Frail, S. R. Kulkarni, A. G. Lyne, *Astrophys. J.* **455**, L55 (1995).
34. A. G. Lyne, R. N. Manchester, J. H. Taylor, *Mon. Not. R. Astron. Soc.* **213**, 613 (1985).
35. $S_{1950,\mathrm{min}}$ is about 8 μJy for $P \sim$ 2- to 4-ms isolated or long orbital period MSPs.
36. F. A. Rasio, D. C. Heggie, *Astrophys. J.* **445**, L133 (1995).
37. A. S. Fruchter, D. R. Stinebring, J. H. Taylor, *Nature* **333**, 237 (1988).
38. A. R. King, M. B. Davies, M. E. Beer, *Mon. Not. R. Astron. Soc.* **345**, 678 (2003).
39. A. Possenti *et al.*, in *Binary Radio Pulsars*, F. Rasio, I. Stairs, Eds. (ASP Conference Series no. 328, ASP, San Francisco, CA, 2005), p. 189.
40. J. Tassoul, *Astrophys. J.* **444**, 338 (1995).
41. F. A. Rasio, S. A. Shapiro, *Astrophys. J.* **377**, 559 (1991).
42. L. L. Smarr, R. Blandford, *Astrophys. J.* **207**, 574 (1976).
43. N. Wex, *Mon. Not. R. Astron. Soc.* **298**, 997 (1998).
44. E. M. Splaver *et al.*, *Astrophys. J.* **581**, 509 (2002).
45. D. J. Nice, E. M. Splaver, I. H. Stairs, in *Binary Radio Pulsars*, F. Rasio, I. Stairs, Eds. (ASP Conference Series no. 328, ASP, San Francisco, CA, 2005), p. 371.
46. J. M. Lattimer, M. Prakash, *Science* **304**, 536 (2004).
47. The Tempo program is available online at http://pulsar.princeton.edu/tempo.

48. We thank F. Rasio, S. Sigurdsson, and M. van Kerkwijk for extremely useful discussions and J. Herrnstein, L. Greenhill, D. Manchester, A. Lyne, and N. D'Amico for providing or aiding with the Parkes data from 1998 and 2000. J.W.T.H. is a Natural Sciences and Engineering Research Council of Canada (NSERC) Post-Graduate Scholarship–Doctoral fellow. I.H.S. holds an NSERC University Faculty Award and is supported by a Discovery grant and University of British Columbia start-up funds. F.C. thanks support from NSF. V.M.K. holds a Canada Research Chair and is supported by an NSERC Discovery Grant and Steacie Fellowship Supplement, by the Fonds Québécois de la recherche sur la nature et les technologies and Canadian Institute for Advanced Research, and by a New Opportunities Grant from the Canada Foundation for Innovation. D.L.K. is a Pappalardo Fellow.

13 December 2004; accepted 4 January 2005
Published online 13 January 2005;
10.1126/science.1108632
Include this information when citing this paper.

Optical Imaging of Neuronal Populations During Decision-Making

K. L. Briggman,[1] H. D. I. Abarbanel,[2,3] W. B. Kristan Jr.[1]*

We investigated decision-making in the leech nervous system by stimulating identical sensory inputs that sometimes elicit crawling and other times swimming. Neuronal populations were monitored with voltage-sensitive dyes after each stimulus. By quantifying the discrimination time of each neuron, we found single neurons that discriminate before the two behaviors are evident. We used principal component analysis and linear discriminant analysis to find populations of neurons that discriminated earlier than any single neuron. The analysis highlighted the neuron cell 208. Hyperpolarizing cell 208 during a stimulus biases the leech to swim; depolarizing it biases the leech to crawl or to delay swimming.

Understanding the mechanisms of behavioral choice would be a major step in bringing together neuroscience, psychology, and ethology (*1*). Research into decision-making has used several different strategies. One very productive approach is to have a primate make a sensory discrimination between very similar stimuli while the activity of neurons in various parts of the nervous system is recorded (*2*–*8*). A second approach uses choice competition: presenting an animal with two stimuli that produce mutually exclusive behaviors (choices), to see which behavior predominates (*9*). This has led to the notion that behavioral choices are hierarchical. The neuronal mechanism originally proposed to underlie behavioral hierarchies was inhibitory interactions among the neurons responsible for triggering the different behaviors (*10*). Later work has found that neurons capable of eliciting one behavior are often activated during other, sometimes conflicting, behaviors (*11*, *12*). Among other things, this observation suggests that individual decision-making neurons can be multiplexed—they contribute to choosing more than one behavior—and that they trigger behaviors by being active with other combinations of neurons.

We used a third approach to study decision-making: choice variability. We presented a nervous system with identical stimuli that repeatedly produce two different, mutually exclusive behaviors with roughly equal probabilities. This approach allowed us to focus on neurons involved in decision-making that are downstream from neurons used to make sensory discriminations.

We used the isolated central nervous system (CNS) of the medicinal leech. Motor neuron activity patterns characteristic of swimming (*13*) and crawling (*14*) can be elicited from isolated preparations by electrically stimulating peripheral nerves. Such sensory stimulation activates mechanosensory neurons in patterns that mimic touching the leech's skin (*15*). Stimulating the same kinds of mechanosensory neurons in different locations on the leech produces characteristic behaviors like swimming or crawling (*16*, *17*). We follow the terminology proposed by Schall (*18*), referring to the different behavioral outputs as choices and the process leading up to a choice as decision-making.

Previously, recording from neurons intracellularly one at a time, then stimulating them to determine their effect on the initiation of behavior, has successfully uncovered interneurons that activate swimming (*19*, *20*), crawling (*21*), and whole-body shortening (*22*). However, to explore how decisions are made by populations of neurons (*11*), we needed to record from many neurons at once (*23*). We therefore used voltage-sensitive dyes (*24*, *25*) that allowed us to record simultaneously from many neurons in a midbody segmental ganglion at a resolution better than 5 mV (*26*).

The leech CNS makes behavioral choices. The isolated leech CNS consists of a nerve cord connecting 21 segmental ganglia plus a head and tail brain (Fig. 1A). This preparation generates motor patterns that are recognizable as behaviors observed in in-

[1]Division of Biological Sciences, [2]Department of Physics, University of California–San Diego, La Jolla, CA 92093–0357, USA. [3]Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, CA 92093–0402, USA.

*To whom correspondence should be addressed. E-mail: wkristan@ucsd.edu

tact leeches, including swimming and crawling. We exposed one ganglion between ganglia 7 (G7) and 10 (G10) for voltage-sensitive dye imaging. We also recorded extracellularly from at least two peripheral nerves [dorsal posterior (DP) nerves] using suction electrodes. These electrodes were used to both stimulate and record. A train of electrical pulses to a DP nerve mimics a touch to the body wall in an intact leech (*15*) and can elicit both swimming and crawling (Fig. 1B). By stimulating DP nerves between G13 and G16, we evoked swimming in about half of the trials (blue) and crawling in the other half (red).

Each trial lasted 60 s, with an intertrial interval of three min. Although the nerve activity was recorded for the entire trial, the ganglion was imaged during only the initial 10 s (Fig. 1B, green bar). We imaged the neurons on the ventral surface of a midbody segmental ganglion (Fig. 1C). In each preparation, we were able to resolve 130 to 150 of the ~160 known neurons (*27*) on the ventral surface. The fluorescence resonance energy transfer (FRET)–based voltage-sensitive dye we used (*24*, *25*) is very sensitive to small membrane potential fluctuations (Fig. 1D). For comparison, we show the raw data from a swim trial and a crawl trial (Fig. 1, E and F). Almost every neuron was activated immediately after each stimulus. There was a clear difference in the activity of many neurons once the motor pattern was apparent in the nerve recordings, usually after ~4 s. These neurons are presumably central pattern generating (CPG) interneurons or motor neurons that generate the swimming or crawling rhythm. We were more interested in activity differences before this time, between the stimulus and 4 s, when the decision between the two behaviors was made.

Discrimination by single neurons. The activity patterns of a subpopulation of single neurons were able to discriminate swimming trials from crawling trials. Of the neurons that responded to the stimulus, we observed four classes of responses: nondiscriminating (ND) cells, early discriminating (ED) cells, late discriminating (LD) cells, and transiently discriminating (TD) cells (Fig. 2A). We quantified the earliest discrimination time of each single cell (t_{SC}) by performing a sliding window analysis of variance (ANOVA) for each cell (Fig. 2B).

Neurons were ordered by their earliest discrimination times (Fig. 2C). The number of cells that discriminated at some point in time (ED, LD, or TD cells) ranged between 50 and 75% across experiments. We also performed an ANOVA on the nerve recordings (*28*) to determine the time (t_{NERVE}) at which we could discriminate the behaviors on the basis of motor neuron activity (Fig. 2, C and D, green line). We were most interested in cells that discriminate before t_{NERVE} because these cells are predictive of the behaviors and are candidate decision-making neurons. A histogram of the earliest t_{SC} for each cell shows that 17 cells (mean ± SD = 19 ± 6 cells; n = 6 preparations) discriminated before t_{NERVE} (Fig. 2D).

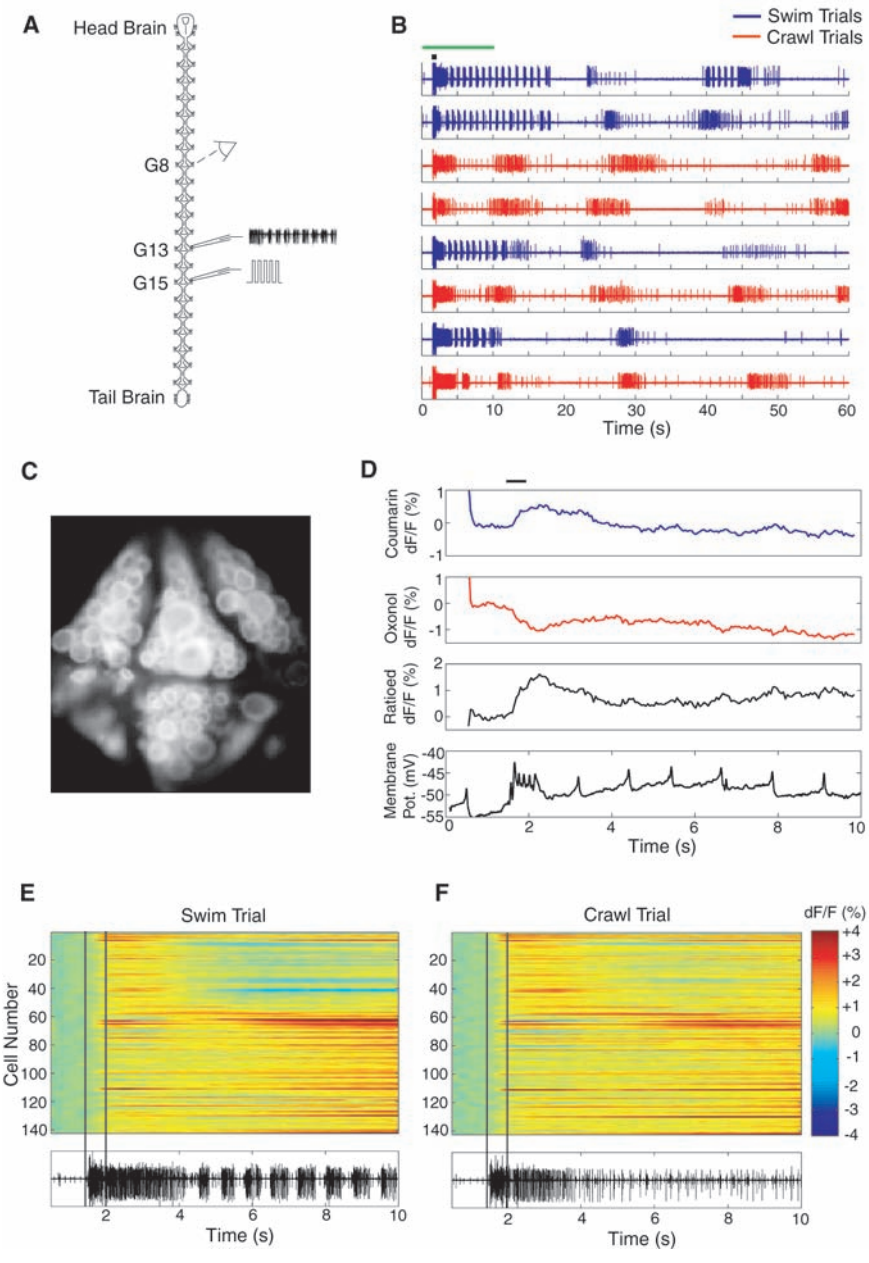


Fig. 1. Recording behavioral choices in the isolated leech CNS. (**A**) Schematic of the sites of recording and stimulation. A midbody (G7 to G10) ganglion was imaged with a voltage-sensitive dye. Suction electrodes were used to record from and stimulate DP nerves (DP13 to DP15). (**B**) Eight sequential 60-s trials demonstrated intertrial variability. The stimulus (2 to 3 V, 10-ms pulses at 15 Hz) lasted 300 ms (small black bar). A ganglion was imaged during the initial 10 s of each trial (green bar). Trials are color-coded by behavior: swimming (blue, ~1- to 2-Hz bursts) or crawling (red, ~0.05- to 0.1-Hz bursts). (**C**) The population of neurons (143 in this example) from which we recorded on the ventral surface of a ganglion. (**D**) We averaged the pixels from each neuron to produce a time-varying record of the percentage change in the fluorescence signal (measured as dF/F). The top two traces are the signals from the two dye molecules, coumarin and oxonol. The largest and least noisy signal is the ratio of these two signals (third trace); it was used for all further analysis. The simultaneous intracellular recording (bottom trace) demonstrates the high sensitivity of the dye (15 to 20% per 100 mV). Pot., potential. (**E** and **F**) Raw optical data from two trials, one which elicited (E) swimming and the other (F) crawling. The optical signal for each neuron is plotted versus time. Color encodes the percent change in fluorescence (dF/F): Positive changes (red) correspond to relative depolarization, and negative changes (blue) correspond to relative hyperpolarization. The panels below each raster plot are simultaneous DP nerve recordings. Vertical black lines indicate the onset and duration of the stimulus.

Discrimination by populations of neurons. We view the nervous system of the leech as a dynamical system (*29*, *30*). Abstractly, the behavioral state of the nervous system at any instant is a point in the phase space of this system (Fig. 3A). Each axis represents a variable that measures the temporal evolution of trajectories in this space (*31*, *32*). For us, these variables are linear combinations of observed neurons. As in the single-cell analysis, we labeled the time at which the trajectories have significantly diverged as the discrimination time. The decision-making period must occur at or before this time. We recognize that the eventual choice may depend on the behavioral state before stimulation, but we did not address this issue in these experiments.

The single-cell discrimination analysis gave us an idea of the number of potential candidate decision-making cells. However, the purpose of simultaneously recording populations of neurons was to look at dynamic interactions among them. Using the multiple recordings, we asked whether a linear combination of neurons could discriminate earlier in time than any single neuron. To address this question, we used two analysis techniques in conjunction: linear discriminant analysis (LDA) (*33*) and principal component analysis (PCA) (*28*, *34*–*36*).

PCA identifies dimensions that separate behaviors. For the experiment shown in Fig. 3C, we performed PCA on a data set containing 143 neurons (dimensions) measured across 14 trials. We show the first three principal components (PCs) (Fig. 3B). The bar graphs represent the dimensionless contribution of each neuron to each PC. Values close to zero indicate a small contribution, and high positive or negative values indicate a large contribution. The first three PCs typically account for 60 to 80% of the overall variance in a data set. The remaining 140 PCs are ignored here for visualization purposes. Effectively, we have reduced the dimensionality of the data set from 143 (neurons) to 3 (linear combinations of neurons, the PCs).

We plotted the data three-dimensionally, using the first three PCs as the axes (Fig. 3C). The trajectories all start in one region and then diverge toward two different regions of the space. From this plot, we extract two features: (i) The separation between the trajectories is an objective measure of decision-making. (ii) The neurons that contribute most to this separation (i.e., have large positive or negative values) are most likely to be responsible for making the choice.

Occasionally, we observed a trial such as the one shown in green in Fig. 3C. This trial would have normally been classified as a crawling trial on the basis of the nerve recording. However, the trajectory in the phase space initially diverges in the direction of the swimming trajectories. After a delay, the trajectory turns and moves toward the direction of the crawling trajectories. Although trials such as this were rare (2 out of 60 trials from six experiments), we interpret this as evidence that decision-making is a dynamical process: The leech nervous system can start to make a decision and subsequently change to an alternative choice.

LDA finds neuronal populations that discriminate. Having reduced the dimensionality of the data set with PCA, we asked (i) at what time did the swimming and crawling trajectories significantly diverge, and (ii) which neurons are responsible for the separation at this time? We thus divided the temporal trajectories into time bins and estimated the linear discriminant for each bin. We then performed an ANOVA on the data in each respective time bin projected onto each linear discriminant. Therefore, each time bin has an associated *P* value. The time at which this *P* value became significant is denoted t_{LDA}. The time bin width and the number of PCs were varied to find the optimal t_{LDA} for each data set (*37*) (fig. S1).

We show an estimated linear discriminant in Fig. 3D using three PCs. Neurons contributing strongly to the linear discriminant were those that best helped discriminate between swimming and crawling at t_{LDA}. In this case, PC3 had the largest weighting, so the linear discriminant direction is similar to PC3.

We wished to visualize the spatial locations of the neurons with large-magnitude contributions to the linear discriminant. We

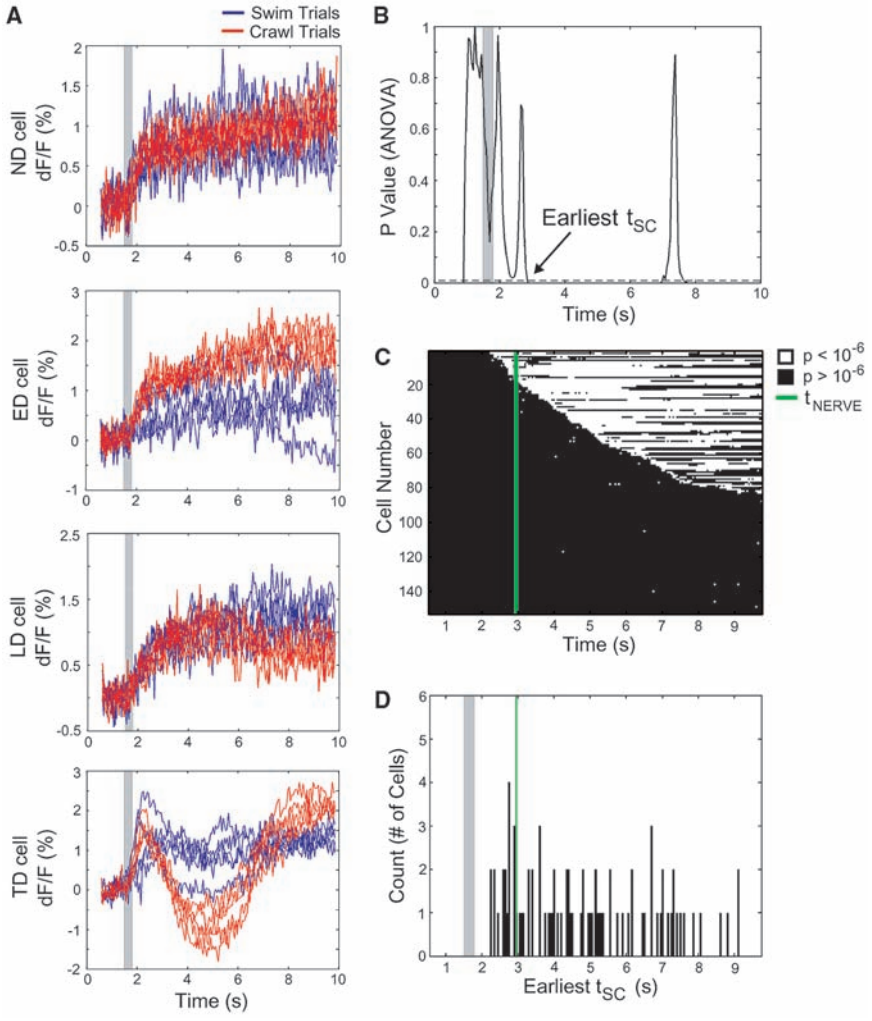


Fig. 2. Single-cell discrimination. (**A**) Examples of the four classes of discrimination responses observed. Each graph plots overlapping raw fluorescence traces from five swimming (blue) and five crawling (red) trials for a single cell. In this and all subsequent plots, the gray shaded region denotes the time and duration of the stimulus. (**B**) A sliding-window ANOVA was used to quantify the discrimination times for each cell. The earliest discrimination time (t_{SC}) was the time at which the swimming and crawling trajectories significantly diverged (black arrow). (**C**) A raster plot of the ANOVA results from all of the cells from one experiment. Black indicates nonsignificant times, and white indicates significantly different times ($P < 10^{-6}$). The discrimination time based on the nerve recording of the behavior is shown (t_{NERVE}, green line). (**D**) A histogram of the earliest t_{SC} from the raster plot in (C). Seventeen neurons produced significantly different trajectories before t_{NERVE}.

used color to encode the magnitude of each neuron and projected the colors onto a map of the ganglion (Fig. 3E). This spatial ganglion map was used as a guide for identifying candidate decision-making neurons.

Discrimination of single neurons versus neuronal populations. We compared the discrimination times of the earliest single cell (t_{ESC}) and the linear discriminant (t_{LDA}) from six experiments (Fig. 4A). In all experiments, t_{LDA} occurred before t_{ESC}. t_{SC} times (black lines) that occurred before the t_{NERVE} times (green lines) are also plotted.

The neuronal populations contributing highly to the linear discriminant were generally different from the neurons with early single-cell discrimination times (Fig. 4, B and C). Single-cell discrimination times (t_{SC}) in Fig. 4C are color-coded, with yellow representing the earliest t_{SC} times and red representing later t_{SC} times. The ganglion maps from all experiments are given in figs. S2 and S3.

Cell 208 biases decisions. The analyses described to this point were performed within 15 min during an ongoing experiment, so the ganglion maps were used to identify candidate decision-making neurons. To test whether these neurons were sufficient to influence decision-making individually, we passed polarizing current into each of them during the nerve shock. We impaled each candidate neuron and injected hyperpolarizing or depolarizing current before and after the nerve shock. None of the neurons with early single-cell discrimination times (Fig. 4C) significantly affected the elicited behaviors (33 neurons from six preparations). When we tested neurons contributing strongly to the linear discriminant (17 neurons from six preparations), we found a neuron, cell 208, that can selectively bias the decision to swim or to crawl (Fig. 4B, arrow, and fig. S3). When depolarized or hyperpolarized alone, this neuron did not initiate any behaviors (*28*). However, when stimulated during a nerve shock, cell 208 biased the decision toward swimming or crawling (Fig. 4D). With cell 208 hyperpolarized, the nerve shock reliably evoked swimming (blue trials); with it depolarized, the nerve shock evoked crawling or delayed swimming (red trials). In five preparations, the correlation between the level of current injection and the observed behavior (Fig. 4E) was significant ($P < 0.01$, Fisher's exact test). We conclude that this neuron plays a role in decision-making. The neuron was labeled with an intracellular dye in all experiments and identified as cell 208 on the basis of its morphology and electrophysiological properties (*38*).

Discussion. One of the central questions we attempted to address is whether decision-making is performed by single neurons or by neuronal populations (*39*). One extreme proposal is multiple competing circuits in which decision-making neurons for one behavior act by inhibiting the other behaviors, so that only one behavior occurs at a time (*9*, *10*). At the other extreme would be the complete sharing of decision-making neurons by two or more behaviors, with the dynamics of the network determining which behavior is chosen. In the leech, we hypothesize a middle ground in which decision-making neurons are partially shared, where the dynamics of neuronal populations can determine choices, but individual neurons in these populations can profoundly influence decision-making. This view is supported by our results that (i)

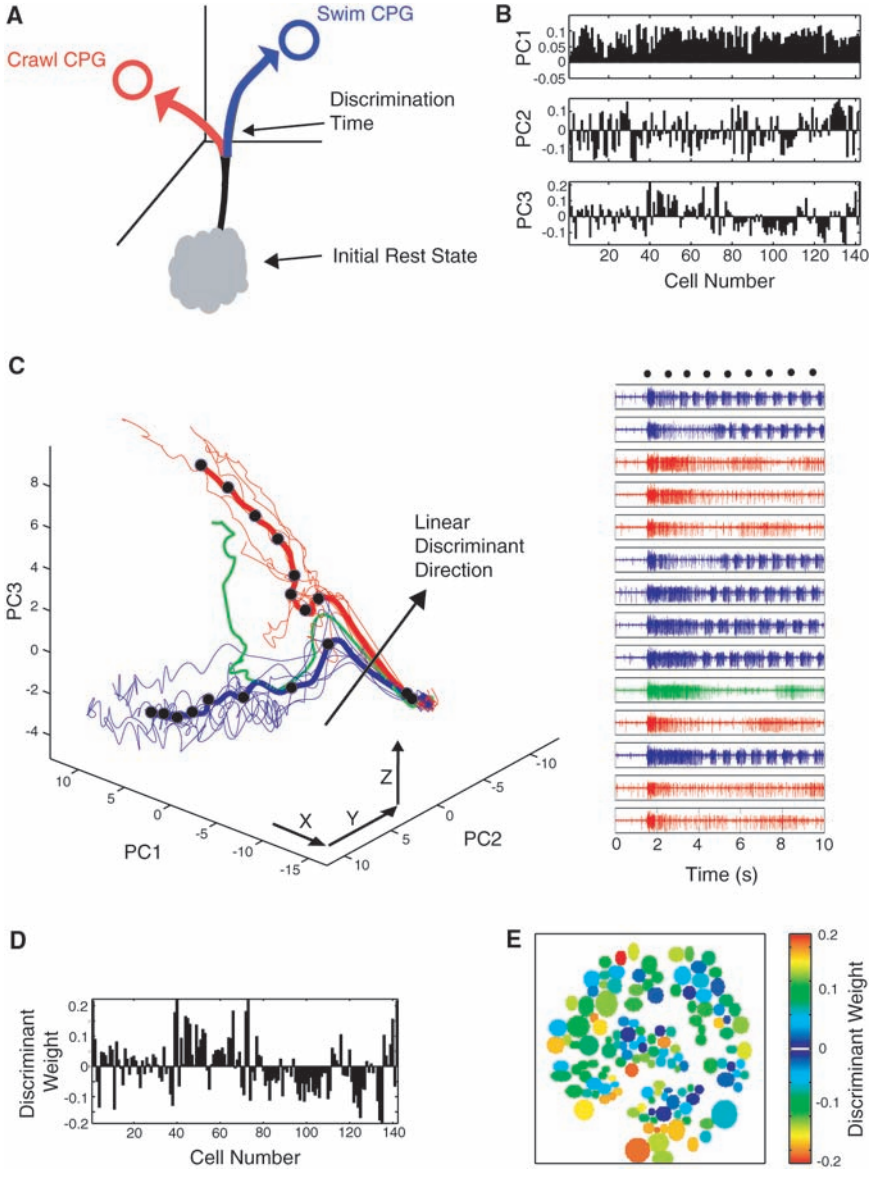


Fig. 3. Population discrimination. (**A**) A schematic of our conceptual framework. The state of the leech CNS at any instant is a point in a *d*-dimensional phase space ($d = 3$ in the drawing). The temporal evolution of its state defines a trajectory in this space. Before stimulation, the state of the nervous system is in a rest region. Upon stimulation, the trajectories diverge toward either a swimming or a crawling region on the basis of which choice was made. (**B**) The first PC from a single experiment. Each PC is a linear combination of the observed neurons in the *N*-dimensional data set ($N = 143$ in this experiment). (**C**) The same data as in (B), projected onto the three PC axes. Each trajectory represents a trial (swimming blue, crawling red). The average swimming and crawling trajectories are in bold. Black dots indicate 1-s intervals along the average trajectories. The right panel plots the simultaneous DP nerve recording for each trial. The black arrow illustrates the direction of a linear discriminant estimated for a 500-ms time bin in the PC space. The direction of the linear discriminant is a weighted vector sum of the three PCs with weights *x*, *y*, and *z*. (**D**) The resulting linear discriminant, a linear combination of the three PCs. This direction is similar to that of PC3 in (B). (**E**) The linear discriminant weightings for each of the neurons projected onto a map of the ganglion. Red indicates large-magnitude (both positive and negative) contributions to the linear discriminant direction; blue indicates small contributions.

neuronal populations discriminated between swimming and crawling earlier in time than did single neurons; (ii) none of the single neurons that discriminated early were able to bias the decision; and (iii) although cell 208 was able to bias the decision, it was part of a population defined by the linear discriminant.

The finding that a linear combination of neurons discriminates early in time shows that there is sufficient information in a population of neurons for an experimenter to predict the ultimate choice. However, this result demonstrates only that the activity pattern of this network correlates with the eventual choice. To show that this information is used during decision-making, we needed to manipulate the network. Ideally, we would like to selectively hyperpolarize or depolarize a population of individual cells; however, we were technologically limited to stimulating one or two cells at a time.

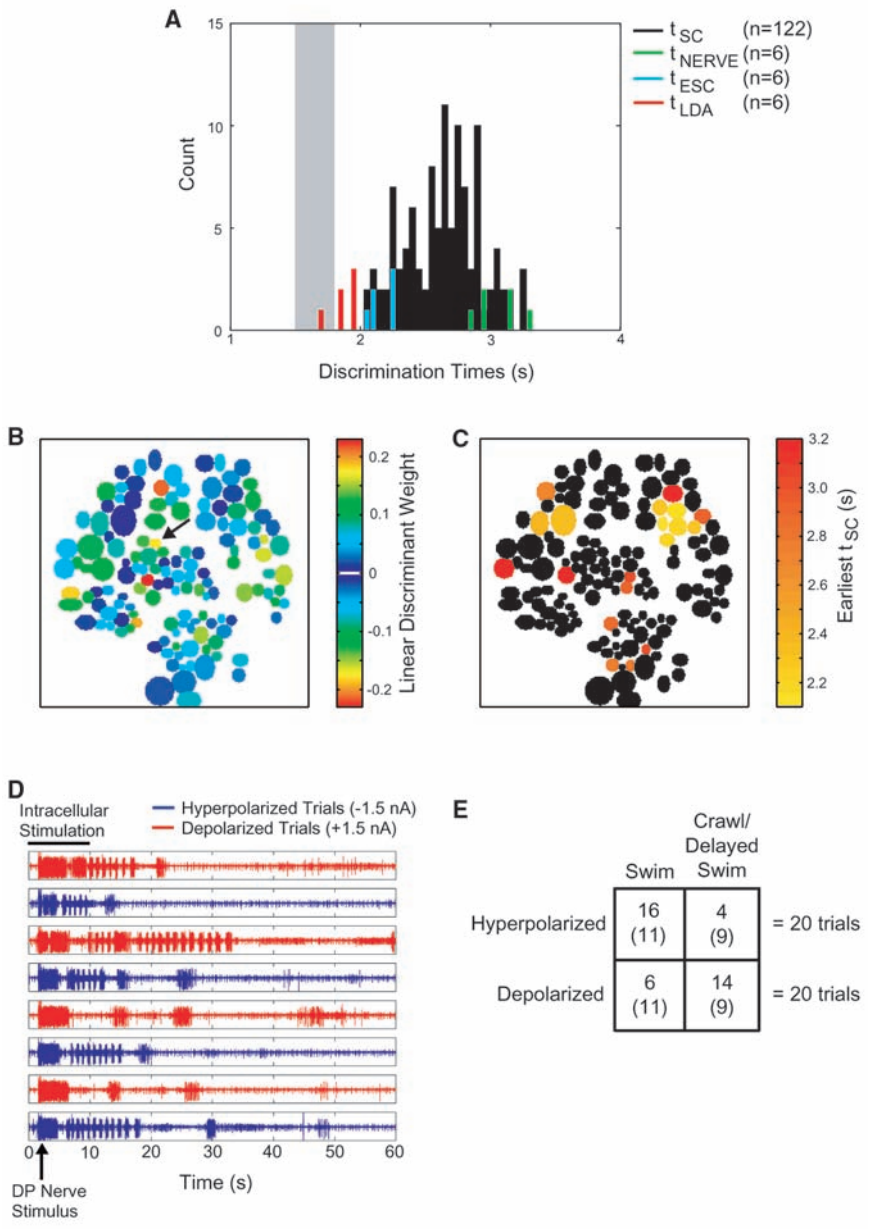


Fig. 4. Comparing single-cell discrimination to population discrimination. (**A**) The distribution of discrimination times from six experiments. Discrimination times for the LDA (t_{LDA}, red), single cells (t_{SC}, black), earliest single cells (t_{ESC}, cyan), and nerve recordings (t_{NERVE}, green) are shown. The average difference between t_{LDA} and t_{ESC} was 290 ± 60 ms (mean ± SEM). Single-cell discrimination times occurring later than the t_{NERVE} for each experiment are omitted. (**B**) The color-coded linear discriminant weightings from a single experiment (color-coding as in Fig. 3E). The black arrow indicates cell 208. (**C**) The color-coded single-cell discrimination times (t_{SC}) from the same experiment. Yellow represents the earliest discrimination times, and red represents later discrimination times. (**D**) The result from one experiment. Cell 208 was intracellularly impaled and alternately depolarized (+1.5 nA) or hyperpolarized (–1.5 nA) for the initial 10 s of each trial (black bar). The membrane resistance of cell 208 at rest was ~50 MΩ. A nerve shock was delivered for 300 ms as in Fig. 1B. In the four trials in which cell 208 was hyperpolarized (blue), the preparation swam within 10 s. In the four remaining trials in which cell 208 was depolarized (red), the preparation either produced the crawling motor pattern or it delayed swimming until the intracellular stimulation ended. (**E**) A contingency table summarizing the pooled results from five preparations. The values in parentheses are the expected counts if the observed behavior did not depend on the stimulus condition.

Although we have not exhaustively tested all candidate neurons detected, most of the neurons tested were not able to bias the behavioral choice. However, we were able to reliably bias the choice with one of the linear discriminant candidates, cell 208. A relatively large depolarization or hyperpolarization of cell 208 biased the choice toward crawling or swimming, respectively. Although this result was statistically significant, cell 208 did not determine the choice on all trials. This less-than-perfect control by cell 208 can probably be credited to the cell 208 homologs in each of the 20 other ganglia (*38*) that we did not stimulate.

Cell 208 has previously been described as a CPG neuron (*38*), but a unique one that connects the swim-initiating network to the swim CPG circuit. It has not been shown to trigger behaviors in intact isolated nerve cords. Our results suggest that it is also part of a decision-making circuit, although we do not yet know its role in this circuit or the mechanism by which it biases the system. Although it is possible that cell 208 is driven by a higher-order decision-making neuron, we have shown that it alone is capable of biasing the entire system.

Most recent research about decision-making has focused on value-based choices in which there is always a right and wrong answer (*2–8*). We propose that the term decision-making should refer to a spectrum of goal-driven behaviors. At the most complex level are conscious, introspective choices that incorporate expected reward (*1*). At the simplest level are reproducible, predictable reflexes. There is a large area of involuntary, subconscious decision-making that has been neglected. For example, suppose your goal is to walk down the street. Which foot do you lead with? This is a choice that is made without conscious effort and will vary from time to time. Although this choice is goal-directed, there is no correct or incorrect choice. In our reduced preparation, there is no obvious value associated with swimming versus crawling, although both choices would achieve the goal of escaping from a stimulus.

Why then does the leech, a relatively simple nervous system, not respond in a predictable manner? We hypothesize two possibilities: Either the choice depends on the rest state before each stimulus, or the state is reset upon stimulation and then diverges stochastically because of noise in the system. The

leech may provide a system in which to resolve this important question.

References and Notes

1. P. W. Glimcher, *Decisions, Uncertainty and the Brain: The Science of Neuroeconomics* (MIT Press, Cambridge, MA, 2003).
2. M. N. Shadlen, W. T. Newsome, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 628 (1996).
3. P. W. Glimcher, *Trends Neurosci.* **24**, 654 (2001).
4. J. D. Schall, *Curr. Biol.* **10**, R404 (2000).
5. M. N. Shadlen, W. T. Newsome, *J. Neurophysiol.* **86**, 1916 (2001).
6. J. I. Gold, M. N. Shadlen, *Nature* **404**, 390 (2000).
7. M. L. Platt, P. W. Glimcher, *Nature* **400**, 233 (1999).
8. R. Romo, E. Salinas, *Annu. Rev. Neurosci.* **24**, 107 (2001).
9. M. P. Kovac, W. J. Davis, *Science* **198**, 632 (1977).
10. M. P. Kovac, W. J. Davis, *J. Neurophysiol.* **43**, 469 (1980).
11. T. Esch, W. B. Kristan Jr., *Integr. Comp. Biol.* **42**, 716 (2002).
12. I. R. Popescu, W. N. Frost, *J. Neurosci.* **22**, 1985 (2002).
13. W. B. Kristan Jr., R. L. Calabrese, *J. Exp. Biol.* **65**, 643 (1976).
14. F. J. Eisenhart, T. W. Cacciatore, W. B. Kristan Jr., *J. Comp. Physiol.* **A186**, 631 (2000).
15. W. B. Kristan Jr., *J. Exp. Biol.* **96**, 161 (1982).
16. W. B. Kristan Jr., S. J. McGirr, G. V. Simpson, *J. Exp. Biol.* **96**, 143 (1982).
17. T. W. Cacciatore, R. Rozenshteyn, W. B. Kristan Jr., *J. Neurosci.* **20**, 1643 (2000).
18. J. D. Schall, *Nature Rev. Neurosci.* **2**, 33 (2001).
19. J. C. Weeks, W. B. Kristan Jr., *J. Exp. Biol.* **77**, 71 (1978).
20. P. D. Brodfuehrer, A. Burns, *Neurobiol. Learn. Mem.* **63**, 192 (1995).
21. T. Esch, K. A. Mesce, W. B. Kristan Jr., *J. Neurosci.* **22**, 11045 (2002).
22. B. K. Shaw, W. B. Kristan Jr., *J. Neurosci.* **17**, 786 (1997).
23. M. Zochowski, L. B. Cohen, G. Fuhrmann, D. Kleinfeld, *J. Neurosci.* **20**, 8485 (2000).
24. J. E. González, K. Oades, Y. Leychkis, A. Harootunian, P. A. Negulescu, *Drug Discov. Today* **4**, 431 (1999).
25. T. W. Cacciatore *et al.*, *Neuron* **23**, 449 (1999).
26. A. L. Taylor, G. W. Cottrell, D. Kleinfeld, W. B. Kristan Jr., *J. Neurosci.* **23**, 11402 (2003).
27. K. J. Muller, J. G. Nicholls, G. S. Stent, *Neurobiology of the Leech* (Cold Spring Harbor, New York, 1981).
28. Materials and methods are available as supporting material on *Science* Online.
29. D. Kaplan, L. Glass, *Understanding Nonlinear Dynamics* (Springer-Verlag, New York, 1995).
30. H. Abarbanel, *Analysis of Observed Chaotic Data* (Springer-Verlag, New York, 1996).
31. D. S. Broomhead, G. P. King, *Phys. D* **20**, 217 (1986).
32. D. S. Broomhead, R. Jones, G. P. King, *J. Phys. A* **20**, 563 (1987).
33. LDA seeks a line in the multidimensional phase space of a system such that grouped data points projected onto the line are maximally separated (i.e., the distributions of swimming versus crawling data points along the line are maximally separated). The slope of this line indicates the relative contribution of each of the variables to this separation. The goal is to find the time at which the swimming and crawling data projected onto a linear discriminant are significantly separated. This technique is susceptible to overfitting when applied in a high-dimensional space with a limited amount of data. To overcome this problem, one may either increase the number of samples or decrease the dimensionality of the system. Phototoxicity limits the amount of data we can collect, so we instead used PCA to reduce the dimensionality of our data sets.
34. PCA rotates the axes of our *N*-dimensional data (where *N* is the number of neurons) so that the first few axes point in the directions of maximal covariance. These new directions are the PCs. A neuron with very different activity between swimming and crawling trials will have a large variance across trials. This neuron would then contribute strongly to one of the first few PCs.
35. R. O. Duda, P. E. Hart, D. G. Stork, *Pattern Classification* (Wiley, New York, ed. 2, 2000).
36. K. Mardia, J. Kent, J. Bibby, *Multivariate Analysis* (Academic Press, San Diego, CA, 1980).
37. The combination of PCA and LDA found linear combinations of neurons that could discriminate the behaviors earlier than any single neuron. Although each of the PCs we discarded did not explain much of the total variance, it is possible that we lost discrimination information by not using them. Thus, although the linear discriminants we found performed well (i.e., discriminated at an early t_{LDA}), an earlier t_{LDA} may have been found had we been able to use all of the PCs. This illustrates the inherent tradeoff between the dimensionality of a system and the amount of data required to adequately sample a high-dimensional space.
38. J. Weeks, *J. Comp. Physiol.* **148**, 265 (1982).
39. W. B. Kristan Jr., B. K. Shaw, *Curr. Opin. Neurobiol.* **7**, 826 (1997).

40. We thank J. E. Gonzalez and R. Y. Tsien for assistance with the FRET voltage-sensitive dyes; Panvera LLC for supplying the dyes gratis; C. Schaffer for assistance with designing the wavelength-switching device; J. Shlens, A. L. Taylor, S. B. Mehta, and E. Thomson for valuable discussions; and H. J. Chiel, M. B. Feller, D. Kleinfeld, and T. J. Sejnowksi for helpful comments and suggestions on an earlier version of this manuscript. Supported by a La Jolla Interfaces in Science Predoctoral Fellowship, funded by the Burroughs Wellcome Fund (K.L.B.); a NSF Integrative Graduate Education and Research Traineeship (IGERT) training grant (K.L.B.); NIH research grants nos. MH43396 (W.B.K.) and NS40110 (H.D.I.A); U.S. Department of Energy grant no. DE-FG03-90ER14138 (H.D.I.A); NSF research grant no. PHY0097134 (H.D.I.A); and the Office of Naval Research, contract no. N00014-00-1-0181 (H.D.I.A).

Supporting Online Material

www.sciencemag.org/cgi/content/full/307/5711/896/DC1
Materials and Methods
SOM Text
Figs. S1 to S3
References and Notes

6 August 2004; accepted 19 November 2004
10.1126/science.1103736

REPORTS

Nodal Quasiparticles and Antinodal Charge Ordering in $Ca_{2-x}Na_xCuO_2Cl_2$

Kyle M. Shen,[1] F. Ronning,[1]* D. H. Lu,[1] F. Baumberger,[1] N. J. C. Ingle,[1] W. S. Lee,[1] W. Meevasana,[1] Y. Kohsaka,[2] M. Azuma,[3] M. Takano,[3] H. Takagi,[2,4] Z.-X. Shen[1]†

Understanding the role of competing states in the cuprates is essential for developing a theory for high-temperature superconductivity. We report angle-resolved photoemission spectroscopy experiments which probe the $4a_0 \times 4a_0$ charge-ordered state discovered by scanning tunneling microscopy in the lightly doped cuprate superconductor $Ca_{2-x}Na_xCuO_2Cl_2$. Our measurements reveal a marked dichotomy between the real- and momentum-space probes, for which charge ordering is emphasized in the tunneling measurements and photoemission is most sensitive to excitations near the node of the *d*-wave superconducting gap. These results emphasize the importance of momentum anisotropy in determining the complex electronic properties of the cuprates and places strong constraints on theoretical models of the charge-ordered state.

To explain the mechanism of high-temperature superconductivity, it is necessary to understand how the parent Mott insulator, characterized by very strong electron-electron repulsion, can evolve into a high–transition temperature (T_c) superconductor upon the addition of a relatively small number of carriers. In the intervening region between the Mott insulator and high-T_c superconductor, the so-called "pseudogap" regime, highly anomalous physical properties have been observed (*1*). Many attempts to explain these unusual properties have centered around the possibility of competing orders, such as orbital currents (*2*), nanoscale charge ordering (*3*, *4*), or electronic phase separation (*5*). The particular importance of charge ordering has recently been underscored by the discovery of a distinct real-space pattern of $4a_0 \times 4a_0$ two-dimensional charge ordering (2DCO) in $Ca_{2-x}Na_xCuO_2Cl_2$ (Na-CCOC) by scanning tunneling microscopy (STM) (*6*). Angle-resolved photoemission spectroscopy (ARPES) can provide crucial

[1]Departments of Applied Physics, Physics, and Stanford Synchrotron Laboratory, Stanford University, Stanford, CA 94305, USA. [2]Department of Advanced Materials Science, University of Tokyo, Kashiwa, Chiba 277–8561, Japan. [3]Institute for Chemical Research, Kyoto University, Uji, Kyoto 611–0011, Japan. [4]RIKEN, The Institute for Physical and Chemical Research, Wako 351–0198, Japan.

*Present address: Los Alamos National Laboratory, Los Alamos, NM 87545, USA.
†To whom correspondence should be addressed. E-mail: zxshen@stanford.edu

leech may provide a system in which to resolve this important question.

References and Notes

1. P. W. Glimcher, *Decisions, Uncertainty and the Brain: The Science of Neuroeconomics* (MIT Press, Cambridge, MA, 2003).
2. M. N. Shadlen, W. T. Newsome, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 628 (1996).
3. P. W. Glimcher, *Trends Neurosci.* **24**, 654 (2001).
4. J. D. Schall, *Curr. Biol.* **10**, R404 (2000).
5. M. N. Shadlen, W. T. Newsome, *J. Neurophysiol.* **86**, 1916 (2001).
6. J. I. Gold, M. N. Shadlen, *Nature* **404**, 390 (2000).
7. M. L. Platt, P. W. Glimcher, *Nature* **400**, 233 (1999).
8. R. Romo, E. Salinas, *Annu. Rev. Neurosci.* **24**, 107 (2001).
9. M. P. Kovac, W. J. Davis, *Science* **198**, 632 (1977).
10. M. P. Kovac, W. J. Davis, *J. Neurophysiol.* **43**, 469 (1980).
11. T. Esch, W. B. Kristan Jr., *Integr. Comp. Biol.* **42**, 716 (2002).
12. I. R. Popescu, W. N. Frost, *J. Neurosci.* **22**, 1985 (2002).
13. W. B. Kristan Jr., R. L. Calabrese, *J. Exp. Biol.* **65**, 643 (1976).
14. F. J. Eisenhart, T. W. Cacciatore, W. B. Kristan Jr., *J. Comp. Physiol.* **A186**, 631 (2000).
15. W. B. Kristan Jr., *J. Exp. Biol.* **96**, 161 (1982).
16. W. B. Kristan Jr., S. J. McGirr, G. V. Simpson, *J. Exp. Biol.* **96**, 143 (1982).
17. T. W. Cacciatore, R. Rozenshteyn, W. B. Kristan Jr., *J. Neurosci.* **20**, 1643 (2000).
18. J. D. Schall, *Nature Rev. Neurosci.* **2**, 33 (2001).
19. J. C. Weeks, W. B. Kristan Jr., *J. Exp. Biol.* **77**, 71 (1978).
20. P. D. Brodfuehrer, A. Burns, *Neurobiol. Learn. Mem.* **63**, 192 (1995).
21. T. Esch, K. A. Mesce, W. B. Kristan Jr., *J. Neurosci.* **22**, 11045 (2002).
22. B. K. Shaw, W. B. Kristan Jr., *J. Neurosci.* **17**, 786 (1997).
23. M. Zochowski, L. B. Cohen, G. Fuhrmann, D. Kleinfeld, *J. Neurosci.* **20**, 8485 (2000).
24. J. E. González, K. Oades, Y. Leychkis, A. Harootunian, P. A. Negulescu, *Drug Discov. Today* **4**, 431 (1999).
25. T. W. Cacciatore *et al.*, *Neuron* **23**, 449 (1999).
26. A. L. Taylor, G. W. Cottrell, D. Kleinfeld, W. B. Kristan Jr., *J. Neurosci.* **23**, 11402 (2003).
27. K. J. Muller, J. G. Nicholls, G. S. Stent, *Neurobiology of the Leech* (Cold Spring Harbor, New York, 1981).
28. Materials and methods are available as supporting material on *Science* Online.
29. D. Kaplan, L. Glass, *Understanding Nonlinear Dynamics* (Springer-Verlag, New York, 1995).
30. H. Abarbanel, *Analysis of Observed Chaotic Data* (Springer-Verlag, New York, 1996).
31. D. S. Broomhead, G. P. King, *Phys. D* **20**, 217 (1986).
32. D. S. Broomhead, R. Jones, G. P. King, *J. Phys. A* **20**, 563 (1987).
33. LDA seeks a line in the multidimensional phase space of a system such that grouped data points projected onto the line are maximally separated (i.e., the distributions of swimming versus crawling data points along the line are maximally separated). The slope of this line indicates the relative contribution of each of the variables to this separation. The goal is to find the time at which the swimming and crawling data projected onto a linear discriminant are significantly separated. This technique is susceptible to overfitting when applied in a high-dimensional space with a limited amount of data. To overcome this problem, one may either increase the number of samples or decrease the dimensionality of the system. Phototoxicity limits the amount of data we can collect, so we instead used PCA to reduce the dimensionality of our data sets.
34. PCA rotates the axes of our *N*-dimensional data (where *N* is the number of neurons) so that the first few axes point in the directions of maximal covariance. These new directions are the PCs. A neuron with very different activity between swimming and crawling trials will have a large variance across trials. This neuron would then contribute strongly to one of the first few PCs.
35. R. O. Duda, P. E. Hart, D. G. Stork, *Pattern Classification* (Wiley, New York, ed. 2, 2000).
36. K. Mardia, J. Kent, J. Bibby, *Multivariate Analysis* (Academic Press, San Diego, CA, 1980).
37. The combination of PCA and LDA found linear combinations of neurons that could discriminate the behaviors earlier than any single neuron. Although each of the PCs we discarded did not explain much of the total variance, it is possible that we lost discrimination information by not using them. Thus, although the linear discriminants we found performed well (i.e., discriminated at an early t_{LDA}), an earlier t_{LDA} may have been found had we been able to use all of the PCs. This illustrates the inherent tradeoff between the dimensionality of a system and the amount of data required to adequately sample a high-dimensional space.
38. J. Weeks, *J. Comp. Physiol.* **148**, 265 (1982).
39. W. B. Kristan Jr., B. K. Shaw, *Curr. Opin. Neurobiol.* **7**, 826 (1997).

40. We thank J. E. Gonzalez and R. Y. Tsien for assistance with the FRET voltage-sensitive dyes; Panvera LLC for supplying the dyes gratis; C. Schaffer for assistance with designing the wavelength-switching device; J. Shlens, A. L. Taylor, S. B. Mehta, and E. Thomson for valuable discussions; and H. J. Chiel, M. B. Feller, D. Kleinfeld, and T. J. Sejnowksi for helpful comments and suggestions on an earlier version of this manuscript. Supported by a La Jolla Interfaces in Science Predoctoral Fellowship, funded by the Burroughs Wellcome Fund (K.L.B.); a NSF Integrative Graduate Education and Research Traineeship (IGERT) training grant (K.L.B.); NIH research grants nos. MH43396 (W.B.K.) and NS40110 (H.D.I.A); U.S. Department of Energy grant no. DE-FG03-90ER14138 (H.D.I.A); NSF research grant no. PHY0097134 (H.D.I.A); and the Office of Naval Research, contract no. N00014-00-1-0181 (H.D.I.A).

Supporting Online Material

www.sciencemag.org/cgi/content/full/307/5711/896/DC1
Materials and Methods
SOM Text
Figs. S1 to S3
References and Notes

6 August 2004; accepted 19 November 2004
10.1126/science.1103736

REPORTS

Nodal Quasiparticles and Antinodal Charge Ordering in $Ca_{2-x}Na_xCuO_2Cl_2$

Kyle M. Shen,[1] F. Ronning,[1]* D. H. Lu,[1] F. Baumberger,[1] N. J. C. Ingle,[1] W. S. Lee,[1] W. Meevasana,[1] Y. Kohsaka,[2] M. Azuma,[3] M. Takano,[3] H. Takagi,[2,4] Z.-X. Shen[1]†

Understanding the role of competing states in the cuprates is essential for developing a theory for high-temperature superconductivity. We report angle-resolved photoemission spectroscopy experiments which probe the $4a_0 \times 4a_0$ charge-ordered state discovered by scanning tunneling microscopy in the lightly doped cuprate superconductor $Ca_{2-x}Na_xCuO_2Cl_2$. Our measurements reveal a marked dichotomy between the real- and momentum-space probes, for which charge ordering is emphasized in the tunneling measurements and photoemission is most sensitive to excitations near the node of the *d*-wave superconducting gap. These results emphasize the importance of momentum anisotropy in determining the complex electronic properties of the cuprates and places strong constraints on theoretical models of the charge-ordered state.

To explain the mechanism of high-temperature superconductivity, it is necessary to understand how the parent Mott insulator, characterized by very strong electron-electron repulsion, can evolve into a high–transition temperature (T_c) superconductor upon the addition of a relatively small number of carriers. In the intervening region between the Mott insulator and high-T_c superconductor, the so-called "pseudogap" regime, highly anomalous physical properties have been observed (*1*). Many attempts to explain these unusual properties have centered around the possibility of competing orders, such as orbital currents (*2*), nanoscale charge ordering (*3*, *4*), or electronic phase separation (*5*). The particular importance of charge ordering has recently been underscored by the discovery of a distinct real-space pattern of $4a_0 \times 4a_0$ two-dimensional charge ordering (2DCO) in $Ca_{2-x}Na_xCuO_2Cl_2$ (Na-CCOC) by scanning tunneling microscopy (STM) (*6*). Angle-resolved photoemission spectroscopy (ARPES) can provide crucial

[1]Departments of Applied Physics, Physics, and Stanford Synchrotron Laboratory, Stanford University, Stanford, CA 94305, USA. [2]Department of Advanced Materials Science, University of Tokyo, Kashiwa, Chiba 277–8561, Japan. [3]Institute for Chemical Research, Kyoto University, Uji, Kyoto 611–0011, Japan. [4]RIKEN, The Institute for Physical and Chemical Research, Wako 351–0198, Japan.

*Present address: Los Alamos National Laboratory, Los Alamos, NM 87545, USA.
†To whom correspondence should be addressed. E-mail: zxshen@stanford.edu

insights into the microscopic nature of this 2DCO and its relationship to the single-particle excitations in **k**-space. We performed ARPES studies of Na-CCOC (x = 0.05, 0.10, and 0.12), allowing us to combine information from the complementary real- and **k**-space electronic probes. Our results reveal a strong momentum anisotropy, in which the 2DCO is associated with strongly suppressed antinodal electronic states that have a nesting wave vector of $|\mathbf{q}| \sim 2\pi/4a_0$, whereas the nodal states dominate the low-energy spectral weight in **k**-space.

ARPES measurements were performed at Beamline 5-4 of the Stanford Synchrotron Radiation Laboratory with the use of single crystals with typical dimensions of 1 × 1 × 0.1 mm grown by a high-pressure flux method (*7*). Na-CCOC is devoid of complications such as superlattice modulations, bilayer splitting, and orthorhombic distortions and is highly 2D with a resistivity anisotropy ρ_c/ρ_{ab} of 10^4 (*8*). The x = 0.10 and 0.12 samples had T_c's of 13 and 22 K, respectively (maximum T_c = 28 K), whereas the x = 0.05 composition was nonsuperconducting. Typical energy and momentum resolutions were 14 meV and 0.35° (corresponding to $\Delta k \sim 0.02\ \pi/a_0$), and samples were measured at pressures lower than 5×10^{-11} torr.

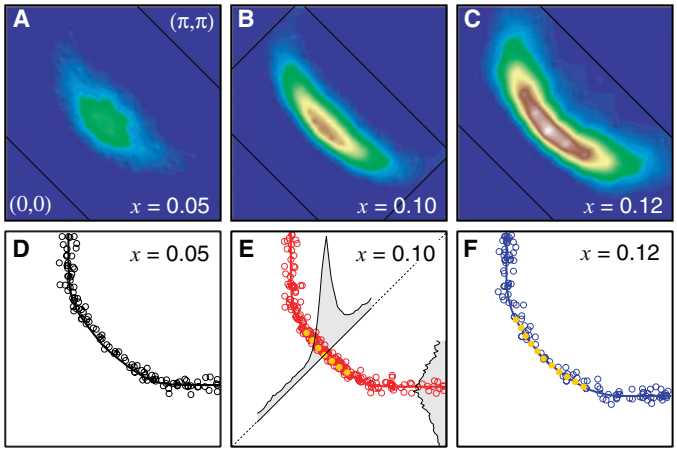


Fig. 1. (**A** to **C**) The momentum distribution of spectral weight within a ±10-meV window around E_F for x = 0.05, 0.10, and 0.12 in one quadrant of the first Brillouin zone. Data were taken at 15 K with $h\upsilon$ = 25.5 eV and a polarization 45° to the Cu-O bond, normalized to a featureless background at high binding energies (–1 eV), and symmetrized along the (0,0)-(π,π) line. The data acquisition range is shown within the black lines. The FS contours shown in (**D** to **F**) were compiled from more than four samples for each composition with photon energies between 16.5 and 28 eV and photon polarizations both parallel to and at 45° to the Cu-O bond direction. Data from these samples constitute the individual points; the best fit is shown as a solid line. The region in which a low-energy peak was typically observed is marked by gold circles. The gray shaded areas in (**E**) represent the momentum distribution of intensity at EF ±10 meV along the (0,0)-(π,π) and (π,0)-(π,π) high-symmetry directions.

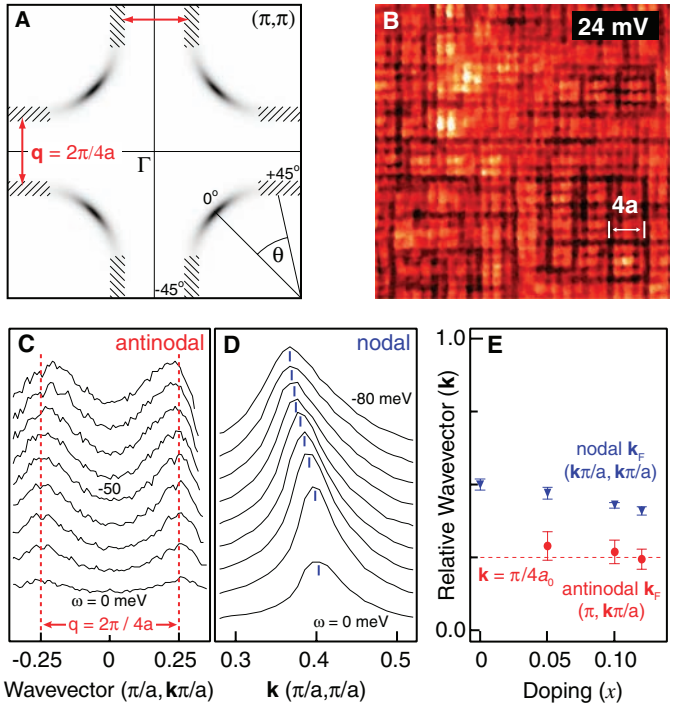


Fig. 2. (**A**) Schematic of the low-lying spectral intensity for x = 0.10. The hatched regions show the nested portions of FS, and the FS angle is defined in the lower right quadrant. (**B**) An STM *dI/dV* map from (*6*) is shown from $Ca_{1.9}Na_{0.1}CuO_2Cl_2$, taken at 24 meV and 100 mK, exhibiting the $4a_0 \times 4a_0$ ordering. MDCs along the antinodal (**C**) and nodal (**D**) directions are shown for $Ca_{1.88}Na_{0.12}CuO_2Cl_2$, taken at 15 K with $h\upsilon$ = 25.5 eV. (**E**) The doping dependence of the $\mathbf{k}_F$ wave vectors along the (0,0)-(π,π) (blue triangles) and (π,0)-(π,π) (red circles) directions. Error bars show the SD from sample to sample.

In Fig. 1, A to C, we show the momentum distribution of spectral weight within a ±10-meV window around the Fermi energy, E_F. The predominance of the nodal states can be seen in the raw data, as the intensity is maximum along the (0,0)-(π,π) nodal direction and drops off rapidly toward (π,0), the antinode. To better quantify the Fermi surface (FS), we have taken the maximal position in each momentum distribution curve (MDC) at E_F, which intersected the FS and identified this as a Fermi wave vector, $\mathbf{k}_F$. To minimize the effects of photoelectron matrix elements or sample-dependent variations, we confirmed our results on additional samples by varying photon energies (between 16.5 and 28 eV) or acquiring data with polarizations parallel to the Cu-O bond direction, or in the second Brillouin zone. All results are summarized in Fig. 1, D to F, and representative MDCs are overlaid in Fig. 1E. Despite the much weaker intensity of the antinodal MDC, its momentum structure nevertheless allows one to define $\mathbf{k}_F$ and establish a continuous contour reminiscent of the predicted noninteracting FS (*9*). Although this approach is robust in extracting the normal-state FS for conventional metallic or even gapped systems, the situation is less clear for strongly correlated systems where the quasiparticle (QP) residue, Z, can be much less than 1. However, we will still refer colloquially to these extracted contours as Fermi surfaces throughout this work (*10*).

The manifestation of the 2DCO in the ARPES spectra can be observed in Fig. 1, D to F, where the weak antinodal segments appear to be well nested and separated by approximately $|\mathbf{q}| \sim 2\pi/4a_0$ (Fig. 2A). In Fig. 2, A and B, we compare a schematic of the low-energy intensity with the real space *dI/dV* map (*6*). This correspondence is exhibited not only in the wave vectors, but also in the unusual energy (ω) dependence of this pattern. The tunneling data exhibit a surprising bias independence (*6*), and our antinodal MDCs (Fig. 2C) also demonstrate a similar insensitivity to ω below 50 meV, in contrast to the dispersive nodal MDCs (Fig. 2D). This unphysical vertical dispersion of the antinodal excitations is highly atypical and almost certainly does not represent the behavior of the actual QP band, as will be discussed later. The doping dependence of the nodal and antinodal $\mathbf{k}_F$'s is summarized in Fig. 2E. The relatively weak doping and ω dependence of the antinodal $\mathbf{k}_F$ is in stark contrast to the expected behavior of a near-E_F van Hove singularity, where both the doping and ω dependence of the MDCs should be sizable. Moreover, the contrast between the strong nodal states and weak antinodal segments is surprising given that the low-energy STM spectra are almost entirely dominated by the commensurate 2DCO (*6*).

This anisotropy can also be observed in the energy distribution curves (EDCs) along the

FS (Fig. 3A). At $x = 0.10$ and 0.12, sharp peaks in the EDCs are observed only near the nodal direction, whereas the antinodes exhibit a depletion of spectral weight over a wide energy range (~200 meV), a phenomenon observed in underdoped cuprates including $Bi_2Sr_2CaCu_2O_{8+\delta}$, $La_{2-x}Sr_xCuO_4$, and Na-CCOC, often referred to as a high-energy pseudogap (*11*). From the EDCs in Fig. 3A, it is clear that the intensity anisotropy in Fig. 1, A to C, is not due simply to the opening of a *d*-wave superconducting (SC) gap, where Bogoliubov QPs should remain well defined over the entire Brillouin zone. In fact, no substantial changes in the spectra were observed upon cooling or warming the samples through T_c (between 10 and 50 K), unlike the cases of near-optimally doped $Bi_2Sr_2CaCu_2O_{8+\delta}$ (*12*) or $YBa_2Cu_3O_{7-\delta}$ (*13*), in which a coherent peak emerges in the antinodal regions below T_c. The absence of any marked change may be associated with the low T_c or the doping level, given that SC antinodal peaks are also absent in underdoped $La_{2-x}Sr_xCuO_4$ and $Bi_2Sr_2CuO_{6+\delta}$ (*11*, *14*, *15*). However, a slight leading-edge gap of the antinodal spectra is observed [8 ± 4 meV for $x = 0.10$, 7 ± 4 meV for $x = 0.12$ (errors are determined from fitting and sample-to-sample variations)], consistent with a low-energy pseudogap. These values are quite consistent with the leading edge gap values observed in compounds with comparable T_c's, such as $La_{2-x}Sr_xCuO_4$ or $Bi_2Sr_2CuO_{6+\delta}$ (*11*, *14*, *15*), whereas they are substantially smaller than in $Bi_2Sr_2CaCu_2O_{8+\delta}$ (*11*), which has a much higher T_c.

In conventional charge-density wave (CDW) systems, nesting of the FS gives rise to a divergent Lindhard susceptibility that results in a Peierls instability. The formation of the CDW results in a larger unit cell and a folding of the original bands, but coherent QPs still exist, as demonstrated by ARPES studies of quasi–2D CDW systems such as $CeTe_3$ (*16*) or In/Cu(001) (*17*). Therefore, the seemingly incoherent antinodal features cannot be produced from FS nesting or ordering alone, so additional factors must be considered. Along these lines, the antinodal spectra from Na-CCOC bear close resemblance to data from $La_{1.2}Sr_{1.8}Mn_2O_7$, in which a similar high-energy pseudogap was observed on portions of the FS that have a nesting wave vector corresponding to CDW ordering (*18*). In this case, the high-energy pseudogap was attributed to the coupling of the electrons to Jahn-Teller distortions of the MnO_6 octahedra, which are believed to play a critical role in the low-energy electronic properties of the manganites, because the Mn^{4+} ions are in a d^3 configuration with unoccupied orbitals of e_g symmetry (*19*). Although the detailed microscopic interactions in the cuprates and manganites are rather different, the two systems may still share a similar general phenomenology.

In Fig. 3B, we show the doping evolution of the low-energy spectral weight associated with the FS and *Z*, where the intensity increases monotonically as a function of doping, behavior unexpected in simple one-electron pictures. This was confirmed by normalizing all MDCs and spectra at all dopings to either the flat high-energy spectra at ~–1 eV or to features in the valence band (–2 to –7 eV), both of which were largely doping independent. Although it appears from Fig. 1, A to C, that the FS elongates toward the antinodes with increasing doping, this effect is illusory and arises simply from the overall increase in intensity, as the shape of the angular distribution of spectral weight does not change appreciably with doping. This suggests that *Z* and the self-energy retain a qualitatively similar **k** anisotropy throughout the underdoped regime.

The absence of sharp antinodal excitations, the doping-dependent growth of low-energy weight, and the **k** anisotropy of FS intensity in Na-CCOC all demonstrate marked resemblances to measurements from lightly doped $La_{2-x}Sr_xCuO_4$ (*14*, *15*). This is shown in Fig. 3, B and C, in which data from $La_{2-x}Sr_xCuO_4$ (*14*) are overlaid onto data from Na-CCOC and scaled to match either the nodal intensity at the same compositions (Fig. 3B) or the general doping trends (Fig. 3C). These similarities suggest an intrinsic commonality between the low-lying excitations across different cuprate families and may imply a generic microscopic origin for these essential nodal states irrespective of other ordering tendencies. At very low doping levels, the nodal excitations should entirely dominate the transport properties, consistent with the high-temperature metallic tendencies observed in very lightly doped cuprates (*20*). Thus any microscopic models of charge ordering must also simultaneously explain and incorporate the existence of coherent nodal states and broad antinodal excitations as essential features of the single-particle excitation spectrum. In addition, similar faint and nested antinodal FS segments have been observed by ARPES in $La_{2-x}Sr_xCuO_4$ (*15*) and could correspond to an analogous 2D charge-ordered state. However, ARPES is a macroscopically averaged probe and cannot distinguish between true 2D ordering and a superposition of 1D domains. Despite the qualitative similarities between the low-energy features of $La_{2-x}Sr_xCuO_4$ and Na-CCOC, there exist important quantitative differences that could potentially be detected by future STM experiments, including the considerably smaller high-energy pseudogap and antinodal wave vector in $La_{2-x}Sr_xCuO_4$ at comparable doping levels (e.g., $|\mathbf{q}| \sim 0.35\ \pi/a_0$ for $x = 0.063$) (*11*, *14*, *15*).

In addition, the identification of similar charge modulation patterns by STM in another compound, $Bi_2Sr_2CaCu_2O_{8+\delta}$, also suggests the possible universality of electronic ordering in all cuprates. Electronic checkerboard patterns are evident in $Bi_2Sr_2CaCu_2O_{8+\delta}$ when SC is destroyed [above T_c (*21*), inside vortex cores (*22*)] or in very underdoped samples (*23*) and may be associated with the absence of sharp antinodal SC excitations in many cuprates studied by ARPES at low doping levels [such as Na-CCOC and $La_{2-x}Sr_xCuO_4$ (*14*, *15*)] or above T_c in underdoped materials [such as $Bi_2Sr_2CaCu_2O_{8+\delta}$ (*11*)]. Although many

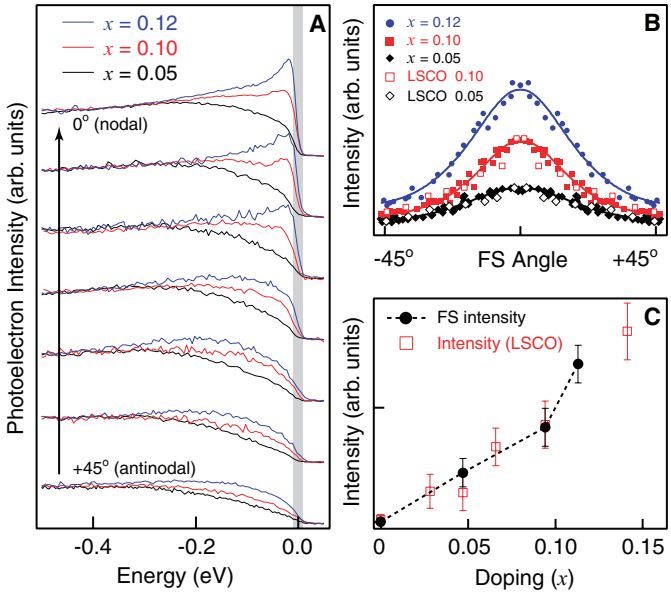


Fig. 3. (**A**) EDCs taken at equal increments along the FS contour from the nodal direction (top) to the antinodal region (bottom) for $x = 0.05$, 0.10, and 0.12. Data was taken at 15 K with $h\upsilon = 25.5$ eV. (**B**) The angular dependence of spectral weight along the FS at $x = 0.05$ (black diamonds), $x = 0.10$ (red squares), and $x = 0.12$ (blue circles), with the angle defined in Fig. 2A, along with data from $La_{2-x}Sr_xCuO_4$ (LSCO) for $x = 0.05$ and $x = 0.10$, shown as open symbols (*14*). (**C**) The doping evolution of the low-lying spectral weight (circles), along with corresponding data from $La_{2-x}Sr_xCuO_4$ (squares) (*14*), with the error bars representing the uncertainty in integrated weight as well as sample-to-sample variations. The data for (B) and (C) were taken within a ±10-meV window, shown in shaded gray in (A). arb. units, arbitrary units.

particular details of the 2DCO state may be material dependent, there appears to be a general correspondence between the broad and nested antinodal FS segments observed by ARPES and the propensity for 2D charge ordering in the lightly doped cuprates.

One recent intriguing model proposed for the 2DCO is a Wigner crystal (WC), whose formation is driven by the Coulomb repulsion between the doped holes (*24*). Although 2D WCs or 1D stripes provide appealing real-space visualizations, addressing the nodal states in a fundamental manner may prove difficult for these models. For the simple WC picture, virtually all holes should be locked into an insulating superlattice and the low-energy QPs arise from the overflow or deficit of holes away from certain favored commensurate dopings. On the other hand, in $La_{2-x}Sr_xCuO_4$ it has been argued that the nodal states may arise from disordered stripes (*25*). In both cases, the nodal states arise more as model-dependent byproducts, not fundamental constituents of the charge-ordered state.

On the other hand, it has long been known that CDW formation and SC are competing instabilities in a wide variety of materials (*26*). This is not entirely surprising, given that the same attractive effective interactions, typically electron-phonon, can give rise to both states in many materials. In the case of Na-CCOC, both *d*-SC and 2DCO appear to compete for the antinodes, and the strength of one order parameter should come at the expense of the other. For instance, although $Ca_{2-x}Na_xCuO_2Cl_2$ is a rather poor high-T_c SC, it exhibits very prominent modulations in the STM *dI*/*dV* maps. On the other hand, $Bi_2Sr_2CaCu_2O_{8+\delta}$ is one of the better high-T_c superconductors (maximum T_c = 96 K), but exhibits far less pronounced charge-density modulations at low energies (*21*–*23*). Along these lines, it is also possible that critical fluctuations between the 2DCO and another ordered state could result in the incoherent antinodal states (*27*), although it is not evident whether the nodal QPs would still remain well defined. A related possibility is that the 2DCO does not represent a CDW of single holes, but instead represents a density wave of preformed *d*-wave Cooper pairs or a pair density wave (PDW) (*28*).

Another potential explanation for the broad antinodal features may come from models based on Franck-Condon broadening, which have been proposed to describe the unusual spectral features in both the lightly doped cuprates (*29*, *30*) and the manganites (*31*), and may therefore explain the similarity of the high-energy pseudogaps found in the both systems. In such a scenario, the strong coupling of the electrons to any bosonic excitations would result in $Z \ll 1$, and spectral weight is transferred to incoherent, multiboson excitations. The high-energy pseudogap behavior in both the cuprates and manganites may then originate from the broad, suppressed tail of incoherent spectral weight, which obscures the very small, but possibly finite coherent QP. The apparently vertical dispersion in the antinodal regime can then be explained as originating not from a real QP band, but rather from this largely incoherent, pseudogapped spectral weight. In this picture, an effective anisotropic coupling could lead to a larger Z (weaker coupling) along the nodal direction and a much smaller, yet still finite Z, at the antinodes (strong coupling). However, this coupling alone may not be sufficient to cause the 2DCO, and it may be the combination of strong coupling and FS nesting which ultimately stabilizes the antinodal charge-ordered state.

Although Na-CCOC may provide the clearest case for relating real-space charge ordering to the anisotropy of electronic states in momentum space, we believe that this is indicative of a generic tendency of the cuprate superconductors. Irrespective of the microscopic model, the dichotomy between the sharp nodal QPs and broad antinodal states suggests the importance of strongly momentum-anisotropic interactions in Na-CCOC and places important restrictions on possible theoretical descriptions of the charge-ordered state and pseudogap phase.

References and Notes

1. T. Timusk, B. Statt, *Rep. Prog. Phys.* **62**, 61 (1999).
2. I. Affleck, J. B. Marston, *Phys. Rev. B* **37**, 3774 (1988).
3. J. Zaanen, O. Gunnarsson, *Phys. Rev. B* **40**, 7391 (1989).
4. J. M. Tranquada *et al.*, *Nature* **375**, 561 (1995).
5. V. J. Emery, S. A. Kivelson, H. Q. Lin, *Phys. Rev. Lett.* **64**, 475 (1990).
6. T. Hanaguri *et al.*, *Nature* **430**, 1001 (2004).
7. Y. Kohsaka *et al.*, *J. Am. Chem. Soc.* **124**, 12275 (2002).
8. K. Waku *et al.*, *Phys. Rev. B* **70**, 134501 (2004).
9. L. F. Mattheiss, *Phys. Rev. B* **42**, 354 (1990).
10. The observed discrepancy between the nominal Na composition and the area enclosed in the contours shown in Fig. 1, C and D, could imply that the experimentally extracted contours may not represent a true FS in the strict Fermi liquid sense, possibly because of the broad antinodal excitations.
11. A. Damascelli, Z. Hussain, Z.-X. Shen, *Rev. Mod. Phys.* **75**, 473 (2003).
12. D. S. Dessau *et al.*, *Phys. Rev. Lett.* **66**, 2160 (1991).
13. D. H. Lu *et al.*, *Phys. Rev. Lett.* **86**, 4370 (2001).
14. T. Yoshida *et al.*, *Phys. Rev. Lett.* **91**, 027001 (2003).
15. X. J. Zhou *et al.*, *Phys. Rev. Lett.* **92**, 187001 (2004).
16. V. Brouet *et al.*, *Phys. Rev. Lett.* **93**, 126405 (2004).
17. T. Nakagawa *et al.*, *Phys. Rev. B* **67**, 241401 (2003).
18. Y.-D. Chuang, A. D. Gromko, D. S. Dessau, T. Kimura, Y. Tokura, *Science* **292**, 1509 (2001).
19. M. B. Salamon, M. Jaime, *Rev. Mod. Phys.* **73**, 583 (2001).
20. Y. Ando, A. N. Lavrov, S. Komiya, K. Segawa, X. F. Sun, *Phys. Rev. Lett.* **87**, 017001 (2001).
21. M. Vershinin *et al.*, *Science* **303**, 1995 (2004).
22. J. E. Hoffman *et al.*, *Science* **295**, 466 (2002).
23. K. McElroy *et al.*, in preparation; preprint available at http://arxiv.org/cond-mat/0404005.
24. H. C. Fu, J. C. Davis, D.-H. Lee, in preparation; preprint available at http://arxiv.org/cond-mat/0403001.
25. M. I. Salkola, V. J. Emery, S. A. Kivelson, *Phys. Rev. Lett.* **77**, 155 (1996).
26. A. M. Gabovich, A. I. Voitenko, J. F. Annett, M. Ausloos, *Supercond. Sci. Technol.* **14**, R1 (2001).
27. S. Sachdev, *Science* **288**, 475 (2000).
28. H.-D. Chen, O. Vafek, A. Yazdani, S.-C. Zhang, *Phys. Rev. Lett.* **93**, 187002 (2004).
29. A. S. Mishchenko, N. Nagaosa, *Phys. Rev. Lett.* **93**, 036402 (2004).
30. K. M. Shen *et al.*, *Phys. Rev. Lett.* **93**, 267002 (2004).
31. V. Perebeinos, P. B. Allen, *Phys. Rev. Lett.* **85**, 5178 (2000).

32. We thank N. P. Armitage, H.-D. Chen, A. Damascelli, J. C. Davis, T. P. Devereaux, H. Eisaki, T. Hanaguri, J. P. Hu, D.-H. Lee, N. Nagaosa, T. Sasagawa, O. Vafek, S. C. Zhang, and X. J. Zhou for enlightening discussions. Stanford Synchrotron Radiation Laboratory is operated by the Department of Energy Office of Basic Energy Science under contract DE-AC03-765F00515. K.M.S. acknowledges Stanford Graduate Fellowships and Natural Sciences and Engineering Research Council of Canada for their support. The ARPES measurements at Stanford were also supported by NSF DMR-0304981 and Office of Naval Research N00014-98-1-0195.

3 August 2004; accepted 30 December 2004
10.1126/science.1103627

Asynchronous Bends in Pacific Seamount Trails: A Case for Extensional Volcanism?

Anthony A. P. Koppers* and Hubert Staudigel

The Gilbert Ridge and Tokelau Seamounts are the only seamount trails in the Pacific Ocean with a sharp 60° bend, similar to the Hawaii-Emperor bend (HEB). These two bends should be coeval with the 47-million-year-old HEB if they were formed by stationary hot spots, and assuming Pacific plate motion only. New $^{40}Ar/^{39}Ar$ ages indicate that the bends in the Gilbert Ridge and Tokelau seamount trail were formed much earlier than the HEB at 67 and 57 million years ago, respectively. Such asynchronous bends cannot be reconciled with the stationary hot spot paradigm, possibly suggesting hot spot motion or magmatism caused by short-term local lithospheric extension.

The Hawaii-Emperor island and seamount chain is the most prominent morphologic feature on the seafloor, with a sharp 60° change in azimuth, called the Hawaii-Emperor bend (HEB). The HEB serves as a textbook example of the fixed hot spot hypothesis, in

particular details of the 2DCO state may be material dependent, there appears to be a general correspondence between the broad and nested antinodal FS segments observed by ARPES and the propensity for 2D charge ordering in the lightly doped cuprates.

One recent intriguing model proposed for the 2DCO is a Wigner crystal (WC), whose formation is driven by the Coulomb repulsion between the doped holes (*24*). Although 2D WCs or 1D stripes provide appealing real-space visualizations, addressing the nodal states in a fundamental manner may prove difficult for these models. For the simple WC picture, virtually all holes should be locked into an insulating superlattice and the low-energy QPs arise from the overflow or deficit of holes away from certain favored commensurate dopings. On the other hand, in $La_{2-x}Sr_xCuO_4$ it has been argued that the nodal states may arise from disordered stripes (*25*). In both cases, the nodal states arise more as model-dependent byproducts, not fundamental constituents of the charge-ordered state.

On the other hand, it has long been known that CDW formation and SC are competing instabilities in a wide variety of materials (*26*). This is not entirely surprising, given that the same attractive effective interactions, typically electron-phonon, can give rise to both states in many materials. In the case of Na-CCOC, both *d*-SC and 2DCO appear to compete for the antinodes, and the strength of one order parameter should come at the expense of the other. For instance, although $Ca_{2-x}Na_xCuO_2Cl_2$ is a rather poor high-T_c SC, it exhibits very prominent modulations in the STM *dI*/*dV* maps. On the other hand, $Bi_2Sr_2CaCu_2O_{8+\delta}$ is one of the better high-T_c superconductors (maximum T_c = 96 K), but exhibits far less pronounced charge-density modulations at low energies (*21*–*23*). Along these lines, it is also possible that critical fluctuations between the 2DCO and another ordered state could result in the incoherent antinodal states (*27*), although it is not evident whether the nodal QPs would still remain well defined. A related possibility is that the 2DCO does not represent a CDW of single holes, but instead represents a density wave of preformed *d*-wave Cooper pairs or a pair density wave (PDW) (*28*).

Another potential explanation for the broad antinodal features may come from models based on Franck-Condon broadening, which have been proposed to describe the unusual spectral features in both the lightly doped cuprates (*29*, *30*) and the manganites (*31*), and may therefore explain the similarity of the high-energy pseudogaps found in the both systems. In such a scenario, the strong coupling of the electrons to any bosonic excitations would result in $Z \ll 1$, and spectral weight is transferred to incoherent, multiboson excitations. The high-energy pseudogap behavior in both the cuprates and manganites may then originate from the broad, suppressed tail of incoherent spectral weight, which obscures the very small, but possibly finite coherent QP. The apparently vertical dispersion in the antinodal regime can then be explained as originating not from a real QP band, but rather from this largely incoherent, pseudogapped spectral weight. In this picture, an effective anisotropic coupling could lead to a larger Z (weaker coupling) along the nodal direction and a much smaller, yet still finite Z, at the antinodes (strong coupling). However, this coupling alone may not be sufficient to cause the 2DCO, and it may be the combination of strong coupling and FS nesting which ultimately stabilizes the antinodal charge-ordered state.

Although Na-CCOC may provide the clearest case for relating real-space charge ordering to the anisotropy of electronic states in momentum space, we believe that this is indicative of a generic tendency of the cuprate superconductors. Irrespective of the microscopic model, the dichotomy between the sharp nodal QPs and broad antinodal states suggests the importance of strongly momentum-anisotropic interactions in Na-CCOC and places important restrictions on possible theoretical descriptions of the charge-ordered state and pseudogap phase.

References and Notes

1. T. Timusk, B. Statt, *Rep. Prog. Phys.* **62**, 61 (1999).
2. I. Affleck, J. B. Marston, *Phys. Rev. B* **37**, 3774 (1988).
3. J. Zaanen, O. Gunnarsson, *Phys. Rev. B* **40**, 7391 (1989).
4. J. M. Tranquada *et al.*, *Nature* **375**, 561 (1995).
5. V. J. Emery, S. A. Kivelson, H. Q. Lin, *Phys. Rev. Lett.* **64**, 475 (1990).
6. T. Hanaguri *et al.*, *Nature* **430**, 1001 (2004).
7. Y. Kohsaka *et al.*, *J. Am. Chem. Soc.* **124**, 12275 (2002).
8. K. Waku *et al.*, *Phys. Rev. B* **70**, 134501 (2004).
9. L. F. Mattheiss, *Phys. Rev. B* **42**, 354 (1990).
10. The observed discrepancy between the nominal Na composition and the area enclosed in the contours shown in Fig. 1, C and D, could imply that the experimentally extracted contours may not represent a true FS in the strict Fermi liquid sense, possibly because of the broad antinodal excitations.
11. A. Damascelli, Z. Hussain, Z.-X. Shen, *Rev. Mod. Phys.* **75**, 473 (2003).
12. D. S. Dessau *et al.*, *Phys. Rev. Lett.* **66**, 2160 (1991).
13. D. H. Lu *et al.*, *Phys. Rev. Lett.* **86**, 4370 (2001).
14. T. Yoshida *et al.*, *Phys. Rev. Lett.* **91**, 027001 (2003).
15. X. J. Zhou *et al.*, *Phys. Rev. Lett.* **92**, 187001 (2004).
16. V. Brouet *et al.*, *Phys. Rev. Lett.* **93**, 126405 (2004).
17. T. Nakagawa *et al.*, *Phys. Rev. B* **67**, 241401 (2003).
18. Y.-D. Chuang, A. D. Gromko, D. S. Dessau, T. Kimura, Y. Tokura, *Science* **292**, 1509 (2001).
19. M. B. Salamon, M. Jaime, *Rev. Mod. Phys.* **73**, 583 (2001).
20. Y. Ando, A. N. Lavrov, S. Komiya, K. Segawa, X. F. Sun, *Phys. Rev. Lett.* **87**, 017001 (2001).
21. M. Vershinin *et al.*, *Science* **303**, 1995 (2004).
22. J. E. Hoffman *et al.*, *Science* **295**, 466 (2002).
23. K. McElroy *et al.*, in preparation; preprint available at http://arxiv.org/cond-mat/0404005.
24. H. C. Fu, J. C. Davis, D.-H. Lee, in preparation; preprint available at http://arxiv.org/cond-mat/0403001.
25. M. I. Salkola, V. J. Emery, S. A. Kivelson, *Phys. Rev. Lett.* **77**, 155 (1996).
26. A. M. Gabovich, A. I. Voitenko, J. F. Annett, M. Ausloos, *Supercond. Sci. Technol.* **14**, R1 (2001).
27. S. Sachdev, *Science* **288**, 475 (2000).
28. H.-D. Chen, O. Vafek, A. Yazdani, S.-C. Zhang, *Phys. Rev. Lett.* **93**, 187002 (2004).
29. A. S. Mishchenko, N. Nagaosa, *Phys. Rev. Lett.* **93**, 036402 (2004).
30. K. M. Shen *et al.*, *Phys. Rev. Lett.* **93**, 267002 (2004).
31. V. Perebeinos, P. B. Allen, *Phys. Rev. Lett.* **85**, 5178 (2000).

32. We thank N. P. Armitage, H.-D. Chen, A. Damascelli, J. C. Davis, T. P. Devereaux, H. Eisaki, T. Hanaguri, J. P. Hu, D.-H. Lee, N. Nagaosa, T. Sasagawa, O. Vafek, S. C. Zhang, and X. J. Zhou for enlightening discussions. Stanford Synchrotron Radiation Laboratory is operated by the Department of Energy Office of Basic Energy Science under contract DE-AC03-765F00515. K.M.S. acknowledges Stanford Graduate Fellowships and Natural Sciences and Engineering Research Council of Canada for their support. The ARPES measurements at Stanford were also supported by NSF DMR-0304981 and Office of Naval Research N00014-98-1-0195.

3 August 2004; accepted 30 December 2004
10.1126/science.1103627

Asynchronous Bends in Pacific Seamount Trails: A Case for Extensional Volcanism?

Anthony A. P. Koppers* and Hubert Staudigel

The Gilbert Ridge and Tokelau Seamounts are the only seamount trails in the Pacific Ocean with a sharp 60° bend, similar to the Hawaii-Emperor bend (HEB). These two bends should be coeval with the 47-million-year-old HEB if they were formed by stationary hot spots, and assuming Pacific plate motion only. New $^{40}Ar/^{39}Ar$ ages indicate that the bends in the Gilbert Ridge and Tokelau seamount trail were formed much earlier than the HEB at 67 and 57 million years ago, respectively. Such asynchronous bends cannot be reconciled with the stationary hot spot paradigm, possibly suggesting hot spot motion or magmatism caused by short-term local lithospheric extension.

The Hawaii-Emperor island and seamount chain is the most prominent morphologic feature on the seafloor, with a sharp 60° change in azimuth, called the Hawaii-Emperor bend (HEB). The HEB serves as a textbook example of the fixed hot spot hypothesis, in

which changes in the azimuth of volcanic lineaments are explained by changes in plate motion, and the hot spots that created these volcanoes remain fixed beneath the moving tectonic plates. Recent suggestions that the HEB was formed by a combination of plate and hot spot motion (*1*, *2*) weaken the fixed hot spot hypothesis, undermining the direct use of seamount trends and the age progressions therein for determining absolute plate-motion vectors. We studied two other HEB-type bends in the Pacific to understand the underlying causes of such azimuthal changes in intraplate volcanic lineaments. These bends can be found at the southern ends of the Gilbert Ridge and Tokelau seamount trail (Fig. 1) that recently were attributed to a group of now-extinct hot spots in the central Pacific Ocean (*3*, *4*), even though volcanism in these chains terminated shortly after the bends formed. The Louisville seamount trail is not useful for an independent age constraint of the HEB, because it shows only a very broad curvature at its bend (*5*).

We explored and dredged the Gilbert Ridge and Tokelau seamount trail aboard the research vessel *R/V Melville* and dated the dredged samples using the $^{40}Ar/^{39}Ar$ dating technique. Our dredging targeted deep volcanic features at the base of these seamounts, and we recovered mostly volcanic materials from which we dated acid-leached groundmass separates (*6*, *7*), or plagioclase, hornblende, and biotite mineral phases (Table 1). The results of our dating study are plotted on the maps in Fig. 2, where we compare the actual $^{40}Ar/^{39}Ar$ age distribution to the predicted plate-motion vectors and seamount ages along the Gilbert Ridge and Tokelau seamount trail from the most recent Pacific plate-motion models (*3*, *4*). These models suggest that the northern and southern parts of the Tokelau seamount trail were formed by two currently inactive hot spots located 5° and 11° to the north of Mcdonald hot spot, whereas the Gilbert Ridge is a composite seamount trail related to three hot spots located 1°, 5°, and 9° north of the Cook-Austral islands. However, our data show that the HEB-type bends in both locations were not formed at 47 million years ago (Ma), as assumed in these models, but rather around 67 Ma for the Gilbert Ridge and 57 Ma for the Tokelau seamount trail. The asynchronous timing of these HEB-type bends causes us to reject the idea that they originated from a group of fixed hot spots and requires us to explore other models for the development of bends in seamount trails. Three alternate models may explain the asynchronous timing of these bends: (i) decelerating motion of hot spot plumes, (ii) channeling of magma away from the hot spots, and (iii) extension of the Pacific plate caused by short-term changes in local plate stresses.

Decelerating motion of hot spots may produce similarly appearing bends asynchronously, in particular, when short-lived mantle plumes terminate magma production and become stationary during their final ascent through a more viscous upper mantle. Hot spot motion is inherent to large-scale mantle convection, where buoyant mantle plumes may shift from fixed and vertically straight positions to tilted positions that are not anchored in the deep mantle (*8*–*10*). The Hawaiian hot spot may have displayed such behavior between 80 and 47 Ma, when it experienced a north-to-south motion at 44 mm/year (*2*, *11*). In this scenario, the NNW Emperor seamount trail azimuth reflects the vector sum of this hot spot motion and a constant plate motion to the NW over the past 80 million years. A considerable slowing of the Hawaiian hot spot around 47 Ma would gradually change the azimuth of its hot spot track toward the northwestern azimuth of Pacific plate motion, as reflected in the Hawaiian seamount trail (*1*, *11*, *12*). Such a deceleration from a south-moving hot spot may also explain the origin of the Gilbert Ridge and Tokelau HEB-type bends, with a termination of hot spot motion around 67 and 57 Ma, respectively. Although this model clearly offers a potential mechanism, there are some problems with it as well. In particular, this model requires two closely spaced small-scale mantle plumes with very consistent hot spot motions for some time, but rapid deceleration and termination at different times, which appear

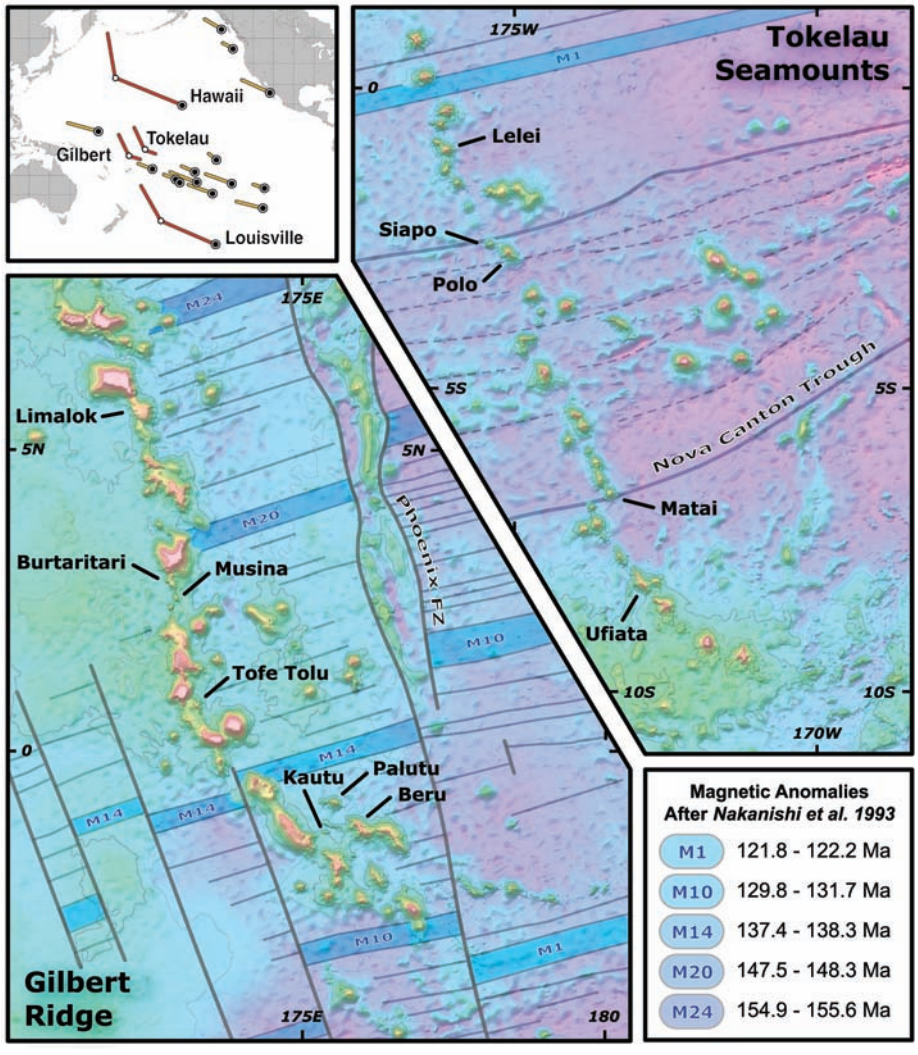


Fig. 1. Bathymetric maps for the Gilbert Ridge and Tokelau seamount trails based on a combination of multibeam data from the AVON Leg 2 cruise onboard the *R/V Melville* and global predicted bathymetry (*17*). Magnetic anomalies and fracture zones in the underlying Pacific oceanic crust (*18*) have been indicated to highlight the lithospheric structure of the Pacific plate. See Table 1 for Web links to the Seamount Catalog for multibeam data, grids, and detailed bathymetric seamount maps.

Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093–0225, USA.

*To whom correspondence should be addressed. E-mail: akoppers@ucsd.edu

unlikely in the same local mantle-convection regime. In addition, the Gilbert Ridge shows ages predominantly in the range between 69 and 65 Ma, suggesting very rapid hot spot motion on the order of 130 mm/year, a velocity that is well beyond the average (~10 mm/year) hot spot motions that are predicted in numerical mantle flow models (*13*).

At first sight, the rather complex structure of the local seafloor (Fig. 1) makes magma diversion away from the main seamount trails a good option for explaining the southeastern deflections observed in the Gilbert Ridge and Tokelau seamount trail. However, a brief examination of the ocean floor structure makes this explanation the least likely among the three options discussed here. Both seamount trails were formed on 50- to 65-million-year-old oceanic crust but terminate in locations of relatively thick and cold lithosphere, just before plate motion would have moved thinner and more vulnerable lithosphere over the mantle plume upwellings. The Gilbert Ridge died out just before northwestern plate motion would decrease the age of the overriding lithosphere at the east side of the Phoenix transform fault, and the Tokelau chain terminated just as the eastern extension of the Manihiki plateau passed the region (Fig. 1). In both cases, we would expect that plume and magma channeling delivered materials very effectively to the thinner Pacific lithosphere, where it met much less resistance to magma transport.

Mantle upwelling caused by plate thinning and extension has been considered as another reason for the development of some seamount trails (*14*, *15*). Plate extension may be related to long-term and plate-wide processes because they are the result of the long-term subduction and ridge-push forces that determine the overall plate-motion vectors (*16*). They may also be short term (10^5 to 10^6 years) and/or localized in particular regions of a tectonic plate, without an apparent impact on the overall long-term plate-motion vector. The latter scenario may be caused by abrupt changes in the subduction zone force balance, because of the arrival of a weak zone that allows the crust to tear, or because of the arrival of an obstacle to subduction, such as a seamount trail or volcanic plateau. As a result, slab-pull forces may change for short durations, causing extensional volcanism at preexisting zones of weakness that are reactivated by minor local changes in plate stress. The stress field will return to its previous state as soon as the subduction zone forces return to their original configuration. We propose that the southwestern Pacific plate experienced two such short-term extensional phases, one around 67 Ma and one around 57 Ma, possibly reactivating the inactive spreading center that formed the Nova Canton Trough and reactivating seafloor preconditioned by small-volume intraplate volcanism that formed close to this ancient Pacific spreading center around 115 Ma and crosses the Phoenix fracture zone to its east. The Tokelau chain also displays a deflection from its NNW azimuth at 3°S, resulting in offshoots that are similar to the bend at its southern termination. This bend may have formed at the same time as the Gilbert Ridge bend, offering independent support for the 67-Ma phase of local plate extension. Overall, plate extension is the strongest alternative among our three options, but there are very few arguments or clues that positively identify any particular explanation.

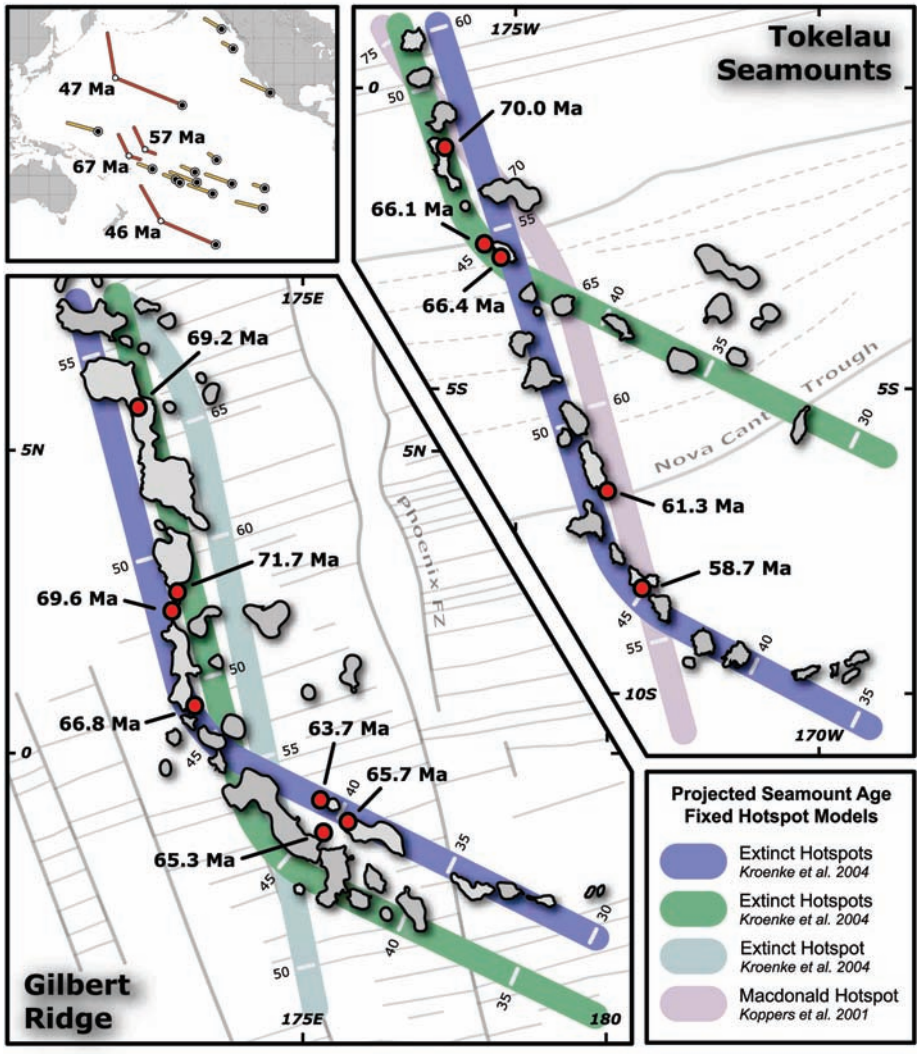


Fig. 2. Average seamount ages based on $^{40}Ar/^{39}Ar$ analyses with respect to the predicted plate motion between 30 and 80 Ma, based on the most recent models from Wessel and Kroenke (*3*, *4*) and assuming fixed hot spots. One prediction by the plate-motion model of Koppers *et al*. 2001 (*19*) for Macdonald hot spot is shown for comparison reasons. All ages are in Ma and are listed in Table 1. The 69.2-Ma age for Limalok guyot was previously published in Koppers *et al*. 2000 (*7*) but has been recalibrated here to reflect the 28.04-Ma Fish Canyon Tuff (FCT-3) biotite age standard.

Our findings influence our views of oceanic intraplate volcanism and absolute Pacific plate motion: (i) The textbook explanation for intraplate volcanism by fixed hot spots is either entirely wrong or insufficient to explain these phenomena. (ii) Hot spots are likely not to be stationary, but move with the convecting mantle. (iii) Non–hot spot/plume models have to be considered for explaining intraplate volcanism, whereby local lithospheric extensions are likely to be an important candidate. (iv) Furthermore, absolute Pacific plate motion, for the time period between 80 and 47 Ma, is extremely poorly constrained. It is not clear if any of the three HEB-type bends on the Pacific plate are caused by a change in plate-motion direction, and it is similarly uncertain if the plate moved NW (along an extended Hawaiian trend) or NNW as indicated by the Emperor seamount trail.

References and Notes

1. I. O. Norton, *Tectonics* **14**, 1080 (1995).
2. J. A. Tarduno *et al.*, *Science* **301**, 1064 (2003).
3. P. Wessel, Y. Harada, L. W. Kroenke, A. Sterling, *Eos* **84**, V32A (2003).

Table 1. Gilbert Ridge and Tokelau Seamounts argon geochronology. Average $^{40}Ar/^{39}Ar$ ages for seamounts in the Gilbert Ridge and Tokelau seamount trails are given. All samples were monitored against Fish Canyon Tuff (FCT-3) biotite (28.04 ± 0.18 Ma) as calibrated by Renne *et al*. (*20*). Reported errors on the $^{40}Ar/^{39}Ar$ ages are on the 95% confidence level, including 0.3 to 0.4% standard deviation on the neutron flux correction factor, or *J* value. All input parameters to the calculations are published in table 2 of Koppers *et al*. (*21*), whereas the ArArCALC v2.2 age calculations are described in Koppers (*22*). Visit the Seamount Catalog at http://earthref.org to find detailed bathymetric maps, grid files, and the multibeam data from the AVON Leg 2 cruise onboard the *R/V Melville* of the Scripps Institution of Oceanography. Use the catalog number in the last column of this table to generate links such as http://earthref.org/cgi-bin/sc.cgi?id=SMNT-067S-1734W that will guide you to the index pages of each dated seamount.

Seamount name	Latitude	Longitude	Age ± 2σ (Ma)	*n*	EarthRef.org Seamount Catalog index number
		Gilbert Ridge			
Burtaritari	2°41.9′N	172°48.2′E	71.7 ± 0.4	2	SMNT-032N-1729E
Musina	2°29.3′N	172°51.7′E	69.6 ± 0.6	2	SMNT-025N-1729E
Tofetolu	0°46.8′N	173°16.1′E	66.8 ± 0.5	2	SMNT-007N-1733E
Palutu	0°55.0′S	175°29.0′E	63.7 ± 0.5	1	SMNT-009S-1755E
Beru	1°08.5′S	175°43.7′E	65.7 ± 0.7	1	SMNT-013S-1760E
Kautu	1°20.8′S	175°20.6′E	65.3 ± 0.5	2	SMNT-014S-1754E
		Tokelau Seamounts			
Lelei	1°04.4′S	176°11.5′	70.0 ± 0.5	3	SMNT-010S-1761W
Siapo	2°36.2′S	175°25.0′	66.1 ± 0.6	4	SMNT-026S-1754W
Polo	2°46.9′S	175°09.0′	66.4 ± 0.6	2	SMNT-027S-1751W
Matai	6°41.9′S	173°28.0′	61.3 ± 0.6	2	SMNT-067S-1734W
Ufiata	8°16.3′S	172°52.7′	58.4 ± 0.3	2	SMNT-082S-1729W

4. L. W. Kroenke, P. Wessel, A. Sterling, in *Origin and Evolution of the Ontong Java Plateau*, J. G. Fitton, J. J. Mahoney, P. J. Wallace, A. D. Saunders, Eds. (Geological Society of London, London, 2004), vol. 229, pp. 9–20.
5. P. Lonsdale, *J. Geophys. Res.* **93**, 3078 (1988).
6. A. A. P. Koppers, H. Staudigel, M. S. Pringle, J. R. Wijbrans, *Geochem. Geophy. Geosyst.* **4**, 1089 (2003).
7. A. A. P. Koppers, H. Staudigel, J. R. Wijbrans, *Chem. Geol.* **166**, 139 (2000).
8. B. Steinberger, R. J. O'Connell, *Geophys. J. Int.* **132**, 412 (1998).
9. B. Steinberger, *J. Geophys. Res.* **105**, 11127 (2000).
10. J. P. Lowman, S. D. King, C. W. Gable, *Geochem. Geophys. Geosyst.* **5**, Q01L01 (2004).
11. R. A. Duncan, R. A. Keller, *Geochem. Geophys. Geosyst.* **5**, Q08L03 (2004).
12. P. Molnar, J. Stock, *Nature* **327**, 587 (1987).
13. B. Steinberger, R. Sutherland, R. J. O'Connell, *Nature* **430**, 167 (2004).
14. E. L. Winterer, D. T. Sandwell, *Nature* **329**, 534 (1987).
15. D. T. Sandwell *et al.*, *J. Geophys. Res.* **100**, 15087 (1995).
16. C. Lithgow-Bertelloni, J. H. Guynn, *J. Geophys. Res.* **109**, B014108 (2004).
17. W. H. F. Smith, D. T. Sandwell, *Science* **277**, 1956 (1997).
18. M. Nakanishi, K. Tamaki, K. Kobayashi, *Geophys. J. Int.* **144**, 535 (1992).
19. A. A. P. Koppers, J. P. Morgan, J. W. Morgan, H. Staudigel, *Earth Planet. Sci. Lett.* **185**, 237 (2001).
20. P. R. Renne *et al.*, *Chem. Geol.* **145**, 117 (1998).
21. A. A. P. Koppers, H. Staudigel, R. A. Duncan, *Geochem. Geophys. Geosyst.* **4**, 8914 (2003).
22. A. A. P. Koppers, *Comput. Geosci.* **28**, 605 (2002).

23. We acknowledge funding from NSF Division of Ocean Sciences through grants NSF–OCE 9730394 and NSF-OCE 0002875. In particular, we express our appreciation of our long-term collaboration with R. A. Duncan and the use of his laboratory. This study was initiated through stimulating brainstorming with J. Phipps Morgan, and the dredging was supported by a group of superb shipmates, including K. Arbesman, J. Dodds, S. W. Herman, J. Konter, T. M. Lassak, J. Phipps Morgan, W. Sager, and R. Taylor. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF.

8 November 2004; accepted 5 January 2005
10.1126/science.1107260

Liquid Carbon, Carbon-Glass Beads, and the Crystallization of Carbon Nanotubes

Walt A. de Heer,[1]* Philippe Poncharal,[2] Claire Berger,[3] Joseph Gezo,[4] Zhimin Song,[1] Jefferson Bettini,[5] Daniel Ugarte[5,6]

The formation of carbon nanotubes in a pure carbon arc in a helium atmosphere is found to involve liquid carbon. Electron microscopy shows a viscous liquid-like amorphous carbon layer covering the surfaces of nanotube-containing millimeter-sized columnar structures from which the cathode deposit is composed. Regularly spaced, submicrometer-sized spherical beads of amorphous carbon are often found on the nanotubes at the surfaces of these columns. Apparently, at the anode, liquid-carbon drops form, which acquire a carbon-glass surface due to rapid evaporative cooling. Nanotubes crystallize inside the supercooled, glass-coated liquid-carbon drops. The carbon-glass layer ultimately coats and beads on the nanotubes near the surface.

The original arc method to produce multiwalled carbon nanotubes (MWNTs) introduced by Ebbesen and Ajayan (*1*, *2*) was an extension of the pure carbon-arc production method for fullerenes developed by Krätschmer and Huffman (*3*). In this method, a pure (catalyst-free) carbon arc (100 A, 30 V) is struck between two carbon electrodes in a helium atmosphere. When the helium pressure is low ($P_{He} \sim 10$ mb), the arc emits a dense fullerene smoke; at higher pressures, carbon nanotubes are formed. In contrast to catalytically produced MWNTs, pure carbon-arc–produced MWNTs are essentially defect free. Here, we present evidence that pure carbon-arc–produced MWNTs form by homogeneous nucleation in liquid carbon, inside elongated carbon droplets coated with a thin layer of carbon glass.

Although the catalytic production process has been extensively studied, little is known about MWNT formation in pure arcs. Earlier proposed formation mechanisms can be classified in two categories. The vapor-growth mechanism involves growth of nanotubes on the cathode in a dense flux of carbon atoms in the arc. Several scenarios have been considered to explain the vapor-solid growth process that is at the same time consistent with the morphology of the nanotube deposits (*4*, *5*). Alternatively, it is known that nanotubes form by heat treatment of various carbonaceous materials (*6*, *7*), and solid-phase production of arc-produced nanotubes was inferred from these observations (*7*). While these mechanisms may indeed produce nanotubes under favorable circumstances, we show that pure carbon-arc–produced nanotubes are formed from a liquid precursor.

Ebbesen and Ajayan (*1*) discovered that when the helium pressure in the Krätschmer-Huffman carbon arc is increased so that $P_{He} >$ 100 mb, a 7-mm-diameter cylindrical carbonaceous deposit forms on the cathode at a rate of about 1 mm per min from a 7-mm-diameter anode. The core of the deposit consists of a basaltic structure of carbon columns parallel to the cylinder axis. The columns are about 1 mm long and about 0.1 mm wide (Fig. 1). They are mechanically stable and easily sepa-

[1]School of Physics, Georgia Institute of Technology, Atlanta, GA 30332, USA. [2]Université Montpellier 2 GDPC, Place Eugène Bataillon 34095 Montpellier, Cedex 5, France. [3]CNRS-LEPES, BP 166, 38042 Grenoble, Cedex 9, France. [4]University of Illinois at Urbana-Champaign, Department of Physics, 1110 West Green Street, Urbana, IL 61801, USA. [5]Laboratório Nacional de Luz Síncrotron, Caixa Postal 6192,13084-971 Campinas SP, Brazil. [6]Instituto de Física Gleb Wataghin, Universidade Estadual de Campinas, Caixa Postal 6165, 13083-970 Campinas SP, Brazil.

*To whom correspondence should be addressed. E-mail: walt.deheer@physics.gatech.edu

Table 1. Gilbert Ridge and Tokelau Seamounts argon geochronology. Average $^{40}Ar/^{39}Ar$ ages for seamounts in the Gilbert Ridge and Tokelau seamount trails are given. All samples were monitored against Fish Canyon Tuff (FCT-3) biotite (28.04 ± 0.18 Ma) as calibrated by Renne *et al*. (*20*). Reported errors on the $^{40}Ar/^{39}Ar$ ages are on the 95% confidence level, including 0.3 to 0.4% standard deviation on the neutron flux correction factor, or *J* value. All input parameters to the calculations are published in table 2 of Koppers *et al*. (*21*), whereas the ArArCALC v2.2 age calculations are described in Koppers (*22*). Visit the Seamount Catalog at http://earthref.org to find detailed bathymetric maps, grid files, and the multibeam data from the AVON Leg 2 cruise onboard the *R/V Melville* of the Scripps Institution of Oceanography. Use the catalog number in the last column of this table to generate links such as http://earthref.org/cgi-bin/sc.cgi?id=SMNT-067S-1734W that will guide you to the index pages of each dated seamount.

Seamount name	Latitude	Longitude	Age ± 2σ (Ma)	*n*	EarthRef.org Seamount Catalog index number
		Gilbert Ridge			
Burtaritari	2°41.9′N	172°48.2′E	71.7 ± 0.4	2	SMNT-032N-1729E
Musina	2°29.3′N	172°51.7′E	69.6 ± 0.6	2	SMNT-025N-1729E
Tofetolu	0°46.8′N	173°16.1′E	66.8 ± 0.5	2	SMNT-007N-1733E
Palutu	0°55.0′S	175°29.0′E	63.7 ± 0.5	1	SMNT-009S-1755E
Beru	1°08.5′S	175°43.7′E	65.7 ± 0.7	1	SMNT-013S-1760E
Kautu	1°20.8′S	175°20.6′E	65.3 ± 0.5	2	SMNT-014S-1754E
		Tokelau Seamounts			
Lelei	1°04.4′S	176°11.5′	70.0 ± 0.5	3	SMNT-010S-1761W
Siapo	2°36.2′S	175°25.0′	66.1 ± 0.6	4	SMNT-026S-1754W
Polo	2°46.9′S	175°09.0′	66.4 ± 0.6	2	SMNT-027S-1751W
Matai	6°41.9′S	173°28.0′	61.3 ± 0.6	2	SMNT-067S-1734W
Ufiata	8°16.3′S	172°52.7′	58.4 ± 0.3	2	SMNT-082S-1729W

4. L. W. Kroenke, P. Wessel, A. Sterling, in *Origin and Evolution of the Ontong Java Plateau*, J. G. Fitton, J. J. Mahoney, P. J. Wallace, A. D. Saunders, Eds. (Geological Society of London, London, 2004), vol. 229, pp. 9–20.
5. P. Lonsdale, *J. Geophys. Res.* **93**, 3078 (1988).
6. A. A. P. Koppers, H. Staudigel, M. S. Pringle, J. R. Wijbrans, *Geochem. Geophy. Geosyst.* **4**, 1089 (2003).
7. A. A. P. Koppers, H. Staudigel, J. R. Wijbrans, *Chem. Geol.* **166**, 139 (2000).
8. B. Steinberger, R. J. O'Connell, *Geophys. J. Int.* **132**, 412 (1998).
9. B. Steinberger, *J. Geophys. Res.* **105**, 11127 (2000).
10. J. P. Lowman, S. D. King, C. W. Gable, *Geochem. Geophys. Geosyst.* **5**, Q01L01 (2004).
11. R. A. Duncan, R. A. Keller, *Geochem. Geophys. Geosyst.* **5**, Q08L03 (2004).
12. P. Molnar, J. Stock, *Nature* **327**, 587 (1987).
13. B. Steinberger, R. Sutherland, R. J. O'Connell, *Nature* **430**, 167 (2004).
14. E. L. Winterer, D. T. Sandwell, *Nature* **329**, 534 (1987).
15. D. T. Sandwell *et al.*, *J. Geophys. Res.* **100**, 15087 (1995).
16. C. Lithgow-Bertelloni, J. H. Guynn, *J. Geophys. Res.* **109**, B014108 (2004).
17. W. H. F. Smith, D. T. Sandwell, *Science* **277**, 1956 (1997).
18. M. Nakanishi, K. Tamaki, K. Kobayashi, *Geophys. J. Int.* **144**, 535 (1992).
19. A. A. P. Koppers, J. P. Morgan, J. W. Morgan, H. Staudigel, *Earth Planet. Sci. Lett.* **185**, 237 (2001).
20. P. R. Renne *et al.*, *Chem. Geol.* **145**, 117 (1998).
21. A. A. P. Koppers, H. Staudigel, R. A. Duncan, *Geochem. Geophys. Geosyst.* **4**, 8914 (2003).
22. A. A. P. Koppers, *Comput. Geosci.* **28**, 605 (2002).

23. We acknowledge funding from NSF Division of Ocean Sciences through grants NSF–OCE 9730394 and NSF-OCE 0002875. In particular, we express our appreciation of our long-term collaboration with R. A. Duncan and the use of his laboratory. This study was initiated through stimulating brainstorming with J. Phipps Morgan, and the dredging was supported by a group of superb shipmates, including K. Arbesman, J. Dodds, S. W. Herman, J. Konter, T. M. Lassak, J. Phipps Morgan, W. Sager, and R. Taylor. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF.

8 November 2004; accepted 5 January 2005
10.1126/science.1107260

Liquid Carbon, Carbon-Glass Beads, and the Crystallization of Carbon Nanotubes

Walt A. de Heer,[1]* Philippe Poncharal,[2] Claire Berger,[3] Joseph Gezo,[4] Zhimin Song,[1] Jefferson Bettini,[5] Daniel Ugarte[5,6]

The formation of carbon nanotubes in a pure carbon arc in a helium atmosphere is found to involve liquid carbon. Electron microscopy shows a viscous liquid-like amorphous carbon layer covering the surfaces of nanotube-containing millimeter-sized columnar structures from which the cathode deposit is composed. Regularly spaced, submicrometer-sized spherical beads of amorphous carbon are often found on the nanotubes at the surfaces of these columns. Apparently, at the anode, liquid-carbon drops form, which acquire a carbon-glass surface due to rapid evaporative cooling. Nanotubes crystallize inside the supercooled, glass-coated liquid-carbon drops. The carbon-glass layer ultimately coats and beads on the nanotubes near the surface.

The original arc method to produce multiwalled carbon nanotubes (MWNTs) introduced by Ebbesen and Ajayan (*1*, *2*) was an extension of the pure carbon-arc production method for fullerenes developed by Krätschmer and Huffman (*3*). In this method, a pure (catalyst-free) carbon arc (100 A, 30 V) is struck between two carbon electrodes in a helium atmosphere. When the helium pressure is low ($P_{He} \sim 10$ mb), the arc emits a dense fullerene smoke; at higher pressures, carbon nanotubes are formed. In contrast to catalytically produced MWNTs, pure carbon-arc–produced MWNTs are essentially defect free. Here, we present evidence that pure carbon-arc–produced MWNTs form by homogeneous nucleation in liquid carbon, inside elongated carbon droplets coated with a thin layer of carbon glass.

Although the catalytic production process has been extensively studied, little is known about MWNT formation in pure arcs. Earlier proposed formation mechanisms can be classified in two categories. The vapor-growth mechanism involves growth of nanotubes on the cathode in a dense flux of carbon atoms in the arc. Several scenarios have been considered to explain the vapor-solid growth process that is at the same time consistent with the morphology of the nanotube deposits (*4*, *5*). Alternatively, it is known that nanotubes form by heat treatment of various carbonaceous materials (*6*, *7*), and solid-phase production of arc-produced nanotubes was inferred from these observations (*7*). While these mechanisms may indeed produce nanotubes under favorable circumstances, we show that pure carbon-arc–produced nanotubes are formed from a liquid precursor.

Ebbesen and Ajayan (*1*) discovered that when the helium pressure in the Krätschmer-Huffman carbon arc is increased so that $P_{He} >$ 100 mb, a 7-mm-diameter cylindrical carbonaceous deposit forms on the cathode at a rate of about 1 mm per min from a 7-mm-diameter anode. The core of the deposit consists of a basaltic structure of carbon columns parallel to the cylinder axis. The columns are about 1 mm long and about 0.1 mm wide (Fig. 1). They are mechanically stable and easily sepa-

[1]School of Physics, Georgia Institute of Technology, Atlanta, GA 30332, USA. [2]Université Montpellier 2 GDPC, Place Eugène Bataillon 34095 Montpellier, Cedex 5, France. [3]CNRS-LEPES, BP 166, 38042 Grenoble, Cedex 9, France. [4]University of Illinois at Urbana-Champaign, Department of Physics, 1110 West Green Street, Urbana, IL 61801, USA. [5]Laboratório Nacional de Luz Síncrotron, Caixa Postal 6192,13084-971 Campinas SP, Brazil. [6]Instituto de Física Gleb Wataghin, Universidade Estadual de Campinas, Caixa Postal 6165, 13083-970 Campinas SP, Brazil.

*To whom correspondence should be addressed. E-mail: walt.deheer@physics.gatech.edu

rated from each other. The interior of a column is composed primarily (~75%) of densely packed, randomly aligned MWNTs, nominally 3 to 20 nm in diameter and several microns in length, and faceted graphitic nanoparticles (~20%); the balance is amorphous carbon (*1*, *2*). The MWNTs are usually structurally quite perfect: Each graphitic layer is continuous and envelops the layer below. They are straight and closed at both ends (*1*, *2*).

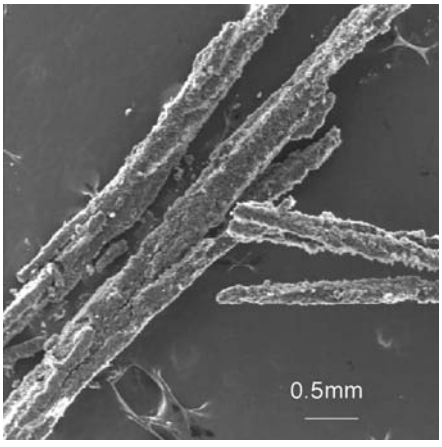


Fig. 1. Several nanotube-containing columns harvested from the interior of the arc deposit. The pure carbon-arc deposits are about 1 cm tall and 7 mm in diameter, with a hard, grayish outer shell and a deep black interior. The columns are closely packed in the deposit, yet easily separated from each other. Beaded nanotubes are occasionally found on the surfaces of these columns. Note, however, that to observe the beads, great care must be taken in separating the columns so as not to break them or destroy their surfaces.

Transmission electron microscopy (TEM) of the columns (*8*) often shows a layer (~100 nm thick) covering their surface, as if the nanotube columns had been dipped in a viscous liquid (Fig. 2). The nanotubes that protrude from the surface of a column are similar in structure to those in the interior; however, we often observe beads on them, as shown in Figs. 2 and 3. TEM and electron diffraction of the beads indicates that the beads are amorphous; energy dispersive x-ray spectroscopy shows that the material is pure carbon. TEM shows that the nanotubes inside the beads are not distorted (Fig. 3). The contrast of the nanotubes inside the beads (Figs. 2 and 3) suggests that the beads have a lower density than the nanotubes. Our measurements of the mass of one of these beads (*9*) assigns a density of 1.4 g/cm^3, which is comparable to that of liquid carbon [1.5 g/cm^3 (*10*)] and less than that of graphite (2.2 g/cm^3). The fact that the beads are sharply imaged, without distortions, indicates that they are not electronically charged by the electron beam, which indicates that the material conducts.

The beads on a specific nanotube are often (but not always) regularly spaced and of similar diameter (Figs. 2 and 3); furthermore, the spacing of the beads is typically comparable to the bead diameter. A clear, extended meniscus is seen at the nanotube-bead interface (Fig. 3), indicating adhesive wetting. These observations indicate that the beads have a liquid-carbon precursor. In particular, Rayleigh showed that a liquid cylinder is unstable: Capillary forces cause the cylinder to break up into approximately equally spaced spherical liquid beads with a spacing that is approximately twice their diameter (*11*). Beading on a liquid-coated fine filament follows the same principle and causes the familiar beading of dewdrops on spider webs (*12*, *13*). Apparently a liquid-carbon layer coated these nanotubes, and this layer subsequently beaded as a result of the Rayleigh instability. Figure 2 shows a remarkable image of an arrested Rayleigh instability. This image suggests that the carbon that coated this nanotube was a viscous liquid. Apparently, cooling caused the viscosity to increase to such a degree that the beading process [whose speed also depends on viscosity (*12*, *13*)] stagnated. This is reminiscent of a glass, because glasses solidify by

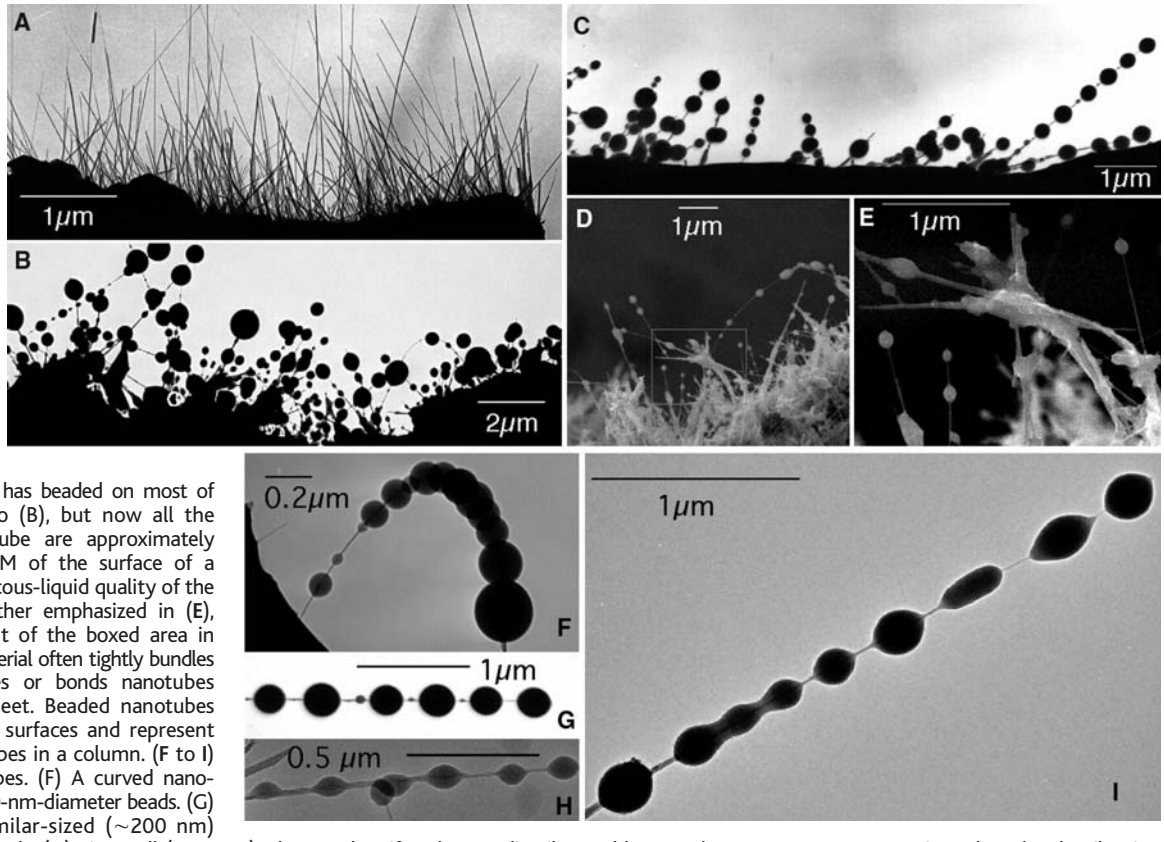


Fig. 2. Electron microscopic analysis of the columns and beaded nanotubes. (**A**) TEM of nanotubes in a column; the interior nanotubes are exposed by breaking a column in two. These nanotubes are straight and relatively free of amorphous carbon particles. They are typical MWNTs: several nm long with diameters from about 5 to 20 nm. (**B**) TEM of the surface of a column, showing nanotubes covered with an amorphous carbon layer that has beaded on most of the tubes. (**C**) Similar to (B), but now all the beads on a specific tube are approximately uniform in size. (**D**) SEM of the surface of a column, showing the viscous-liquid quality of the amorphous coating, further emphasized in (**E**), which is an enlargement of the boxed area in (D). The amorphous material often tightly bundles two or more nanotubes or bonds nanotubes together where they meet. Beaded nanotubes appear only on column surfaces and represent $<10^{-4}$ all of the nanotubes in a column. (**F** to **I**) TEM of beaded nanotubes. (F) A curved nanotube with 12 100- to 200-nm-diameter beads. (G) A nanotube with six similar-sized (~200 nm) and uniformly spaced beads. (H) Six small (~70 nm), elongated, uniformly spaced beads. The nanotube can be seen traversing the beads; the meniscus at the nanotube interface can also be discerned. (I) Arrested beading. The undulations on the second bead are caused by the Rayleigh instability, which ordinarily would cause the structure to separate into three beads. Likewise, surface tension would ordinarily have caused the elongated fifth bead to become more spherical. In this rare example, the beading was arrested because of the cooling of the carbon glass.

becoming more viscous as they cool, without crystallizing and with little change in density (*14*). This assessment is supported by the observation that the beads are formed only at the column surfaces, where cooling rates are expected to be very high (see below). Therefore, we refer to this amorphous material as carbon glass. [Vitreous carbon or glassy carbon, on the other hand, is a low-density amorphous form of carbon typically formed by pyrolysis of resins at lower temperatures; it may be structurally similar to our material (*15*).]

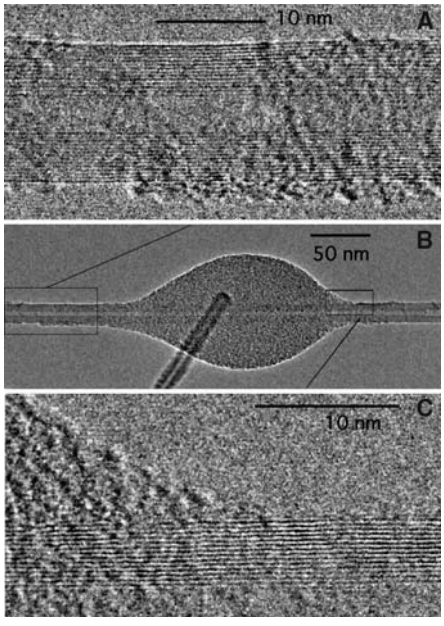


Fig. 3. TEM of a bead on a 15-layer MWNT (a second nanotube is in the background). (**A**) High-resolution image of the furthest extent of the meniscus on the left side of the bead, which here is about 1.3 nm thick. (**B**) Low-resolution image of the small, elongated bead. Note its cylindrical symmetry, the absence of facets, and a very well-formed meniscus. The nanotube is clearly visible inside the bead. (**C**) High-resolution image of the bead meniscus to the right. There is no evidence for graphitic layering in the bead, nor does the bead distort the nanotube.

The nanotubes do not show any signs of damage, nor have they evaporated or dissolved to any degree in the beads. This means that the balls were liquid-like, while the nanotubes were stable, implying bead formation temperatures well below the carbon melting temperature. Looking carefully at the interface between the amorphous beads and the MWNTs, we observe neither extended nor nanometric graphitic pieces, which would be easily detectable in the atomic-resolution electron-microscopy images [see, for example, studies on the formation of graphitic onions by electron-microscopy images (*16*)]. The fact that the thinnest region of the wetting meniscus (Fig. 3) does not reveal signs of even partial epitaxial graphitization also indicates that the liquid was relatively cool when it was deposited on the nanotube. The appearance of the carbon-glass coating indicates that it was deposited on the nanotubes after they were fully formed (Figs. 2 and 3). From this evidence, we reconstruct a probable nanotube formation scenario.

The pure carbon anode is locally heated by electron bombardment from the cathode, which is about 1 mm away, causing the surface to locally liquefy and liquid-carbon globules (whose shape is not known but whose size is related to that of the final columns in the deposit) to be ejected from the anode. Initially, because of the very high vapor pressure of liquid carbon, the surface of a globule will evaporatively cool until the carbon vapor pressure at the surface equals the helium pressure. This occurs at temperature T_B such that the carbon vapor pressure at that temperature equals the helium pressure: $P_C(T_B) = P_{He}$. The initial, very efficient evaporative cooling of the surface (probably distinctive to liquid-carbon surfaces, due to its very large triple-point pressure) causes the globule surface to glassify and possibly seals it. Once the carbon vapor pressure becomes smaller than the helium background pressure, the evaporation rates diminish, and evaporative cooling essentially ceases. [For low P_{He}, evaporation continues, which explains why no cathode deposit is formed (*1*–*3*)]. The more gradual cooling of the interior of the globule causes the liquid carbon to supercool so that carbon crystals (nanotubes and small graphitic particles) homogeneously nucleate and grow (*17*). While the interior liquid transforms to higher density graphitic structures, the hot viscous glass skin adheres to the nanotubes. Subsequently, capillary forces cause the glass coating to bead on the nanotubes. Finally, the hot, relatively plastic globules are deposited on the cathode to form the basaltic structure of aligned columns on the cathode. The process is schematically depicted in Fig. 4.

The triple-point liquid carbon has recently been found to occur at the carbon melting temperature T_m 4800 K, with a carbon vapor pressure $P_C(T_m)$ = 110 bar and liquid density ρ = 1.5 g/cm^3 [determined from average densities of quenched liquid carbon (*9*)]. If the liquid globule surface temperature T_S is such that $P_C(T_S) > P_{He}$, then the surface evaporates at a rate proportional to $P_C(T_S)$: At P_C = 0.5 bar,

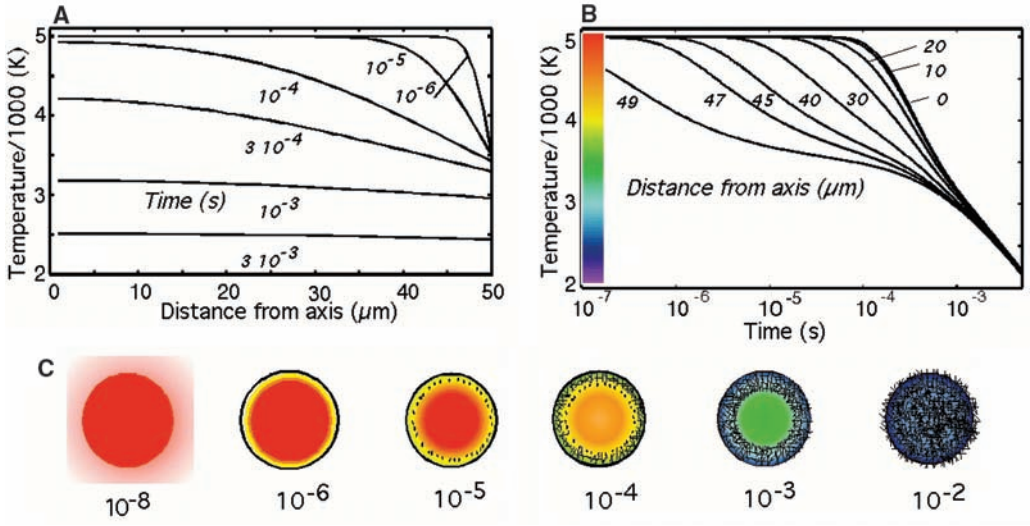


Fig. 4. Temperature evolution of a 50-μm-radius liquid-carbon drop in cold He gas at a pressure of P_{He} = 0.5 bar. T = 5000 K ($>T_m$) at time 0; physical parameters used in the model are from (*6*). (**A**) Temperature as a function of distance from the axis at various times. (**B**) Temperature as a function of time at various locations in the drop. Note that within a few ns, the surface temperature drops to T_B = 3500 K, where $P_C(T_B) = P_{He}$. The very large temperature gradients near the surface, combined with evaporative cooling, cause cooling rates on the order of 10^8 K/s and cause the liquid near the surface to quench to a glass. The interior cools more slowly (on the order of 10^6 K/s), becomes supercooled, and then crystallizes by homogeneous nucleation (*10*). (**C**) Schematic of the cooling of a liquid drop, which starts as a liquid cylindrical drop at time t = 0 at T = 5000; the surface evaporates (upper left). After about 10^{-6} s, the surface has cooled to 3500 K and glassifies. Upon further cooling, nucleation sets in near the surface in the critically supersaturated liquid (t = 10^{-5} s); nanotubes begin their growth as small carbon crystallite seeds. Nanotube growth is conjectured to proceed from these seeds, following the growth front (the interface between the critically supersaturated liquid and the subcritical liquid) from the surface toward the middle as the drop cools. The growth process continues for about 10^{-2} s, filling the drop with MWNTs. Finally, the thin, hot, viscous glass layer adheres to the nanotubes, causing the beading. Larger droplets behave similarly as far as processes near the surfaces are concerned; the overall cooling rates of the interior are approximately inversely proportional to the drop diameter.

the evaporation rate is about 1 monolayer in 10^{-9} s (*18*, *19*), thereby rapidly cooling the surface to T_B, where $P_C(T_B) = P_{He}$. Heat arriving at the surface from the still-hot interior is rapidly dissipated by evaporation. (This is similar to the pinning of the boiling temperature by the ambient pressure.) Hence, initially, the surface temperature is maintained at T_B. For $P_{He} = 0.5$ bar, $T_B \sim 3500$ K. When the globule has lost enough heat (after ~10 μs), the surface temperature reduces below T_B and nonevaporative cooling processes (initially radiation, later convection and conduction) take over. The entire globule then uniformly cools relatively slowly. After ~1 ms, it will have cooled to ~3000 K (Fig. 4). Assuming that the deposit is formed from a stream of D = 100 μm diameter spherical droplets that are charged to the anode potential and accelerated by the electric fields, we estimate the flight time t to be of the order of 1 ms (t is proportional to D).

To get an idea of the temperatures and time scales involved, we considered a 100-μm-diameter liquid drop initially at T = 5000 K (i.e., slightly above T_m) (shown in Fig. 4). (Cylinders and spheres give rather similar results.) Within 10^{-8} s, the surface temperature reduces to $T_B \sim 3500$ K and stays there for about 100 μs while the interior cools. After about 100 μs, the temperature gradients near the surface reduce to about 10 K/μm, which is low enough for the drop surface to slowly and uniformly cool below T_B. It is important to note that T_B = 3500 K corresponds to about 0.72 T_m, which is below the minimum temperature T_{sc} typical for the onset of homogeneous nucleation in small drops of pure elements [i.e., $T_{sc} = 0.80 \pm 0.07\ T_m$ (*17*)]. Hence, because the liquid in the interior of the drop is supercooled, homogeneous nucleation and crystal growth proceeds, and the liberated heat is dissipated at the surface by evaporation: The drop is a thermostat, whose temperature is fixed by P_{He}. In the initial evaporative cooling phase (which last about 10 μs), maximum cooling rates are calculated to exceed 10^9 K/s within the first micron from the drop surface, while cooling rates are greatly reduced deeper inside the drop. Note that high cooling rates, from above to well below the melting temperature, are required to quench liquids into glasses (*14*).

We demonstrated above that conditions in liquid-carbon drops are favorable for homogeneous nucleation. The question remains why most of the liquid carbon crystallizes predominantly in MWNTs, whereas a smaller fraction crystallizes in faceted onion-like particles. The microscopic evidence does not provide an answer to this question. We may, however, hypothesize that elongated crystal growth is promoted by the radial temperature gradient in the drop during the cooling. As indicated in Fig. 4, the nanotubes start as small crystallites near the surface at the interface of the critically supersaturated liquid and the subcritical liquid. As the drop cools, this interface propagates toward the center, and the crystal growth follows this growth front, promoting the growth of elongated crystals. [It is also notable that the formation of small, elongated graphitic structures is favored over quasispherical structures even in isotropic thermal conditions (*6*).] To describe accurately the nanotube crystallization process will require much more refined considerations. Here, we primarily provide information on the environment in which the nanotubes grow.

The most important features of our observations are the role of liquid carbon in the nanotube formation and the role of the helium pressure to establish the temperature of the globules when the nanotubes are formed, which turns out to be in the range expected for homogeneous nucleation in supercooled drops. Hence, these nanotubes (and graphite nanoparticles) are, in fact, nanocrystallites grown in small, supercooled liquid globules in the usual way (*17*). The formation of a thin, amorphous carbon skin occurs at the globule surfaces, where the thermal quenching rates are very high. This material is liquid on well-formed nanotubes without dissolving or otherwise distorting the nanotubes. Furthermore, the material has a density that is close to that of liquid carbon. Both the processes leading to the amorphous material and the properties that we can ascribe to it are characteristic of a glass: It is almost certain that this material is in fact carbon glass.

References and Notes

1. T. W. Ebbesen, P. M. Ajayan, *Nature* **358**, 220 (1992).
2. P. M. Ajayan, T. W. Ebbesen, *Rep. Prog. Phys.* **60**, 1025 (1997).
3. W. Krätschmer, L. D. Lamb, K. Fostiropoulos, D. R. Huffman, *Nature* **347**, 354 (1990).
4. D. T. Colbert *et al.*, *Science* **266**, 1218 (1994).
5. E. G. Gamaly, T. W. Ebbesen, *Phys. Rev. B* **52**, 2083 (1995).
6. W. A. de Heer, D. Ugarte, *Chem. Phys. Lett.* **207**, 480 (1993).
7. P. J. F. Harris, S. C. Tsang, J. B. Claridge, M. L. H. Green, *J. Chem. Soc. Faraday Trans. 1* **90**, 2799 (1994).
8. The nanotube deposits were generated in an arc (20 V, 100 A) between a 7-mm-diameter graphite anode and a 2.5-cm-diameter graphite cathode in a 500 mb He atmosphere. The samples were analyzed with scanning microscopy (JSM 6330F) and high-resolution TEM (JEM-3010 URP, 0.17-nm point resolution) at Campinas, Brazil; low-resolution TEM images were analyzed with a JEOL 100CX II at Georgia Tech. Individual nanotube-containing carbon columns were glued with silver paint on conventional TEM copper grids and suspended over the grid holes. It must be emphasized that the sample must be prepared with extreme care to minimize the manipulation of the individual carbon columns and conserve the pristine surface where the beaded NTs are present.
9. P. Poncharal, Z. L. Wang, D. Ugarte, W. A. de Heer, *Science* **283**, 1513 (1999).
10. The physical constants for liquid carbon are from (*20*).
11. L. Rayleigh, *Proc. R. Soc. A* **8**, 425 (1879).
12. S. L. Goren, *J. Fluid Mech.* **137**, 363 (1962).
13. I. L. Kliakhandler, S. H. Davis, S. G. Bankoff, *J. Fluid Mech.* **429**, 381 (2001).
14. C. A. Angell, *J. Phys. Chem. Solids* **49**, 863 (1988).
15. B. B. O'Malley, I. Snook, D. McCulloch, *Phys. Rev. B* **57**, 14148 (1998).
16. D. Ugarte, *Nature* **359**, 707 (1992).
17. J. C. Brice, *Crystal Growth Processes* (Blackie, London, 1986).
18. The evaporation rate is estimated from the vapor pressure by noting that, in thermodynamic equilibrium with the vapor (and a unit sticking coefficient), the rate of atoms arriving on the surface is balanced by an equal rate of evaporated atoms (*21*).
19. W. A. de Heer, *Rev. Mod. Phys.* **65**, 611 (1993).
20. M. Musella, C. Ronchi, M. Brykin, M. Sheindlin, *J. Appl. Phys.* **84**, 2530 (1998).
21. C. E. Klots, *Z. Phys. D* **5**, 83 (1987).

22. This work was supported by NSF grants 9971412 and 0404084; we acknowledge P. C. Silva, the Georgia Institute of Technology Electron Microscopy Facility, and particularly Z. L. Wang for assistance.

2 November 2004; accepted 28 December 2004
10.1126/science.1107035

Independent Origins of Middle Ear Bones in Monotremes and Therians

Thomas H. Rich,[1,2]* James A. Hopson,[3] Anne M. Musser,[4] Timothy F. Flannery,[5] Patricia Vickers-Rich[2]

A dentary of the oldest known monotreme, the Early Cretaceous *Teinolophos trusleri*, has an internal mandibular trough, which in outgroups to mammals houses accessory jaw bones, and probable contact facets for angular, coronoid, and splenial bones. Certain of these accessory bones were detached from the mandible to become middle ear bones in mammals. Evidence that the angular (homologous with the mammalian ectotympanic) and the articular and prearticular (homologous with the mammalian malleus) bones retained attachment to the lower jaw in a basal monotreme indicates that the definitive mammalian middle ear evolved independently in living monotremes and therians (marsupials and placentals).

In the evolutionary transition from primitive synapsids (the so-called mammal-like reptiles) to extant mammals, the dentary bone of the lower jaw established a neomorphic articulation with the squamosal bone of the skull, and three of the accessory lower jaw bones

the evaporation rate is about 1 monolayer in 10^{-9} s (*18*, *19*), thereby rapidly cooling the surface to T_B, where $P_C(T_B) = P_{He}$. Heat arriving at the surface from the still-hot interior is rapidly dissipated by evaporation. (This is similar to the pinning of the boiling temperature by the ambient pressure.) Hence, initially, the surface temperature is maintained at T_B. For $P_{He} = 0.5$ bar, $T_B \sim 3500$ K. When the globule has lost enough heat (after ~10 μs), the surface temperature reduces below T_B and nonevaporative cooling processes (initially radiation, later convection and conduction) take over. The entire globule then uniformly cools relatively slowly. After ~1 ms, it will have cooled to ~3000 K (Fig. 4). Assuming that the deposit is formed from a stream of D = 100 μm diameter spherical droplets that are charged to the anode potential and accelerated by the electric fields, we estimate the flight time t to be of the order of 1 ms (t is proportional to D).

To get an idea of the temperatures and time scales involved, we considered a 100-μm-diameter liquid drop initially at T = 5000 K (i.e., slightly above T_m) (shown in Fig. 4). (Cylinders and spheres give rather similar results.) Within 10^{-8} s, the surface temperature reduces to $T_B \sim 3500$ K and stays there for about 100 μs while the interior cools. After about 100 μs, the temperature gradients near the surface reduce to about 10 K/μm, which is low enough for the drop surface to slowly and uniformly cool below T_B. It is important to note that T_B = 3500 K corresponds to about 0.72 T_m, which is below the minimum temperature T_{sc} typical for the onset of homogeneous nucleation in small drops of pure elements [i.e., $T_{sc} = 0.80 \pm 0.07\ T_m$ (*17*)]. Hence, because the liquid in the interior of the drop is supercooled, homogeneous nucleation and crystal growth proceeds, and the liberated heat is dissipated at the surface by evaporation: The drop is a thermostat, whose temperature is fixed by P_{He}. In the initial evaporative cooling phase (which last about 10 μs), maximum cooling rates are calculated to exceed 10^9 K/s within the first micron from the drop surface, while cooling rates are greatly reduced deeper inside the drop. Note that high cooling rates, from above to well below the melting temperature, are required to quench liquids into glasses (*14*).

We demonstrated above that conditions in liquid-carbon drops are favorable for homogeneous nucleation. The question remains why most of the liquid carbon crystallizes predominantly in MWNTs, whereas a smaller fraction crystallizes in faceted onion-like particles. The microscopic evidence does not provide an answer to this question. We may, however, hypothesize that elongated crystal growth is promoted by the radial temperature gradient in the drop during the cooling. As indicated in Fig. 4, the nanotubes start as small crystallites near the surface at the interface of the critically supersaturated liquid and the subcritical liquid. As the drop cools, this interface propagates toward the center, and the crystal growth follows this growth front, promoting the growth of elongated crystals. [It is also notable that the formation of small, elongated graphitic structures is favored over quasispherical structures even in isotropic thermal conditions (*6*).] To describe accurately the nanotube crystallization process will require much more refined considerations. Here, we primarily provide information on the environment in which the nanotubes grow.

The most important features of our observations are the role of liquid carbon in the nanotube formation and the role of the helium pressure to establish the temperature of the globules when the nanotubes are formed, which turns out to be in the range expected for homogeneous nucleation in supercooled drops. Hence, these nanotubes (and graphite nanoparticles) are, in fact, nanocrystallites grown in small, supercooled liquid globules in the usual way (*17*). The formation of a thin, amorphous carbon skin occurs at the globule surfaces, where the thermal quenching rates are very high. This material is liquid on well-formed nanotubes without dissolving or otherwise distorting the nanotubes. Furthermore, the material has a density that is close to that of liquid carbon. Both the processes leading to the amorphous material and the properties that we can ascribe to it are characteristic of a glass: It is almost certain that this material is in fact carbon glass.

References and Notes

1. T. W. Ebbesen, P. M. Ajayan, *Nature* **358**, 220 (1992).
2. P. M. Ajayan, T. W. Ebbesen, *Rep. Prog. Phys.* **60**, 1025 (1997).
3. W. Krätschmer, L. D. Lamb, K. Fostiropoulos, D. R. Huffman, *Nature* **347**, 354 (1990).
4. D. T. Colbert *et al.*, *Science* **266**, 1218 (1994).
5. E. G. Gamaly, T. W. Ebbesen, *Phys. Rev. B* **52**, 2083 (1995).
6. W. A. de Heer, D. Ugarte, *Chem. Phys. Lett.* **207**, 480 (1993).
7. P. J. F. Harris, S. C. Tsang, J. B. Claridge, M. L. H. Green, *J. Chem. Soc. Faraday Trans. 1* **90**, 2799 (1994).
8. The nanotube deposits were generated in an arc (20 V, 100 A) between a 7-mm-diameter graphite anode and a 2.5-cm-diameter graphite cathode in a 500 mb He atmosphere. The samples were analyzed with scanning microscopy (JSM 6330F) and high-resolution TEM (JEM-3010 URP, 0.17-nm point resolution) at Campinas, Brazil; low-resolution TEM images were analyzed with a JEOL 100CX II at Georgia Tech. Individual nanotube-containing carbon columns were glued with silver paint on conventional TEM copper grids and suspended over the grid holes. It must be emphasized that the sample must be prepared with extreme care to minimize the manipulation of the individual carbon columns and conserve the pristine surface where the beaded NTs are present.
9. P. Poncharal, Z. L. Wang, D. Ugarte, W. A. de Heer, *Science* **283**, 1513 (1999).
10. The physical constants for liquid carbon are from (*20*).
11. L. Rayleigh, *Proc. R. Soc. A* **8**, 425 (1879).
12. S. L. Goren, *J. Fluid Mech.* **137**, 363 (1962).
13. I. L. Kliakhandler, S. H. Davis, S. G. Bankoff, *J. Fluid Mech.* **429**, 381 (2001).
14. C. A. Angell, *J. Phys. Chem. Solids* **49**, 863 (1988).
15. B. B. O'Malley, I. Snook, D. McCulloch, *Phys. Rev. B* **57**, 14148 (1998).
16. D. Ugarte, *Nature* **359**, 707 (1992).
17. J. C. Brice, *Crystal Growth Processes* (Blackie, London, 1986).
18. The evaporation rate is estimated from the vapor pressure by noting that, in thermodynamic equilibrium with the vapor (and a unit sticking coefficient), the rate of atoms arriving on the surface is balanced by an equal rate of evaporated atoms (*21*).
19. W. A. de Heer, *Rev. Mod. Phys.* **65**, 611 (1993).
20. M. Musella, C. Ronchi, M. Brykin, M. Sheindlin, *J. Appl. Phys.* **84**, 2530 (1998).
21. C. E. Klots, *Z. Phys. D* **5**, 83 (1987).

22. This work was supported by NSF grants 9971412 and 0404084; we acknowledge P. C. Silva, the Georgia Institute of Technology Electron Microscopy Facility, and particularly Z. L. Wang for assistance.

2 November 2004; accepted 28 December 2004
10.1126/science.1107035

Independent Origins of Middle Ear Bones in Monotremes and Therians

Thomas H. Rich,[1,2]* James A. Hopson,[3] Anne M. Musser,[4] Timothy F. Flannery,[5] Patricia Vickers-Rich[2]

A dentary of the oldest known monotreme, the Early Cretaceous *Teinolophos trusleri*, has an internal mandibular trough, which in outgroups to mammals houses accessory jaw bones, and probable contact facets for angular, coronoid, and splenial bones. Certain of these accessory bones were detached from the mandible to become middle ear bones in mammals. Evidence that the angular (homologous with the mammalian ectotympanic) and the articular and prearticular (homologous with the mammalian malleus) bones retained attachment to the lower jaw in a basal monotreme indicates that the definitive mammalian middle ear evolved independently in living monotremes and therians (marsupials and placentals).

In the evolutionary transition from primitive synapsids (the so-called mammal-like reptiles) to extant mammals, the dentary bone of the lower jaw established a neomorphic articulation with the squamosal bone of the skull, and three of the accessory lower jaw bones

were transformed to sound-transmitting elements of the middle ear (*1*). The articular and prearticular bones became the mammalian malleus, and the angular bone became the ectotympanic (or tympanic ring), which supports the eardrum (Fig. 1A). An additional synapsid element, the quadrate (which with the articular forms the primitive synapsid jaw joint), became the mammalian incus. A controversy exists as to whether the transformation of jaw bones to middle ear bones occurred independently in the two clades of living mammals: the Monotremata (platypuses and echidnas) and Theria (marsupials and placentals). In other words, did the accessory jaw bones that gave rise to the ear ossicles and ectotympanic become detached from the lower jaw only once (*2*–*7*) in the common ancestry of monotremes and therians (a monophyletic origin), or did they become detached from the jaw independently in the two living groups subsequent to their evolutionary divergence from a common ancestor (a polyphyletic origin) (*1*, *8*–*10*)? Assertions of fundamental differences in development and morphology between monotreme and therian ears are no longer supported (*4*, *11*), so the primary argument for a polyphyletic origin lies in the existence of mammal-like dentaries from the Late Triassic to Early Cretaceous (*12*–*16*) that show evidence of a persisting contact of putative ear bone homologs with the lower jaw. Unfortunately, the contentious nature of the phylogenetic relations of Mesozoic mammals has until now prevented the establishment of a reliable link between fossil mammals with accessory jaw bones and living and fossil mammals with true ear ossicles. Here we present evidence of such a link between a fossil monotreme with accessory jaw bones (Fig. 2) and living monotremes in which certain of those bones are entirely within the middle ear.

The transformation of jaw bones to ear bones was first documented embryologically, when it was recognized that the reptilian articular bone and the greater part of the mammalian malleus were both formed by ossification of the posterior portion of Meckel's cartilage, the embryonic cartilage of the lower jaw (*17*) (Fig. 1B). With the establishment of an embryonic (or neonate) contact of the mammalian dentary bone with the squamosal of the skull, the middle portion of Meckel's cartilage atrophies, leaving the incus (= reptilian quadrate), malleus (= articular + prearticular), and ectotympanic (= angular) ligamentously attached to the ear region of the skull (*1*).

One of the most compelling pieces of fossil evidence for the transformation of jaw to ear bones is seen in *Morganucodon* (Fig. 1C), a Late Triassic/Early Jurassic near-mammal (mammaliaform), in which the lower jaw consists of a greatly enlarged dentary and six highly reduced accessory bones, with the primitive articular-quadrate jaw joint lying adjacent to a newly evolved mammalian dentary-squamosal jaw joint (*18*). The rod of accessory jaw bones that includes the homologs of the ear bones of living mammals lies in a prominent trough on the posteromedial side of the dentary (Fig. 3A). It is the presence of such a trough in the dentaries of presumed mammals [such as *Shuotherium* (*13*) and *Asfaltomylos* (*16*)] thought to be closely allied either to monotremes (*19*) or to therians (*20*, *21*) (Fig. 4) that provides the principal evidence for a possible multiple origin of the definitive mammalian ear bones. However, the presence of a mandibular trough in an Early Cretaceous monotreme provides what we consider to be the most compelling evidence for the polyphyletic origin of the definitive mammalian middle ear.

Teinolophos trusleri (*22*–*24*) is known from six mandibular fragments, of which the most informative are the holotype (Museum Victoria specimen number NMV P208231) (Fig. 2, A and B) and the most complete referred dentary (NMV P212933) (Fig. 2, C to H). The latter is in much better condition than the holotype, which is crushed and distorted (*22*, *23*), and preserves fine surface detail absent in the type. All were found at the Early Cretaceous (Early Aptian) Flat Rocks fossil vertebrate locality [38°39′40″ ± 02″S, 145°40′52″ ± 03″ E (World Geodetic Standard 1984)] in the Wonthaggi Formation on the shore platform of the Bunurong Marine Park, Victoria, Australia (*22*).

A striking feature of the medial surface of the referred dentary of *T. trusleri* is the prominent mandibular trough (m.t, Figs. 2D and 3C) that extends forward from the notch above the angular process (a.p) to the large mandibular foramen (m.f, damaged because of a V-shaped break). A similar trough in *Morganucodon* (Fig. 3A), docodontids (Fig. 3B) (*25*, *26*), and other basal mammaliaforms houses a rod of accessory jaw bones including the articular (that is, the ossified posterior end of Meckel's cartilage) and three dermal bones, the dorsal and lateral surangular, the ventral and lateral angular, and the medial prearticular (Fig. 1C). Together the three dermal elements form a hollow cylinder that presumably enclosed a persistent anterior extension of Meckel's carti-

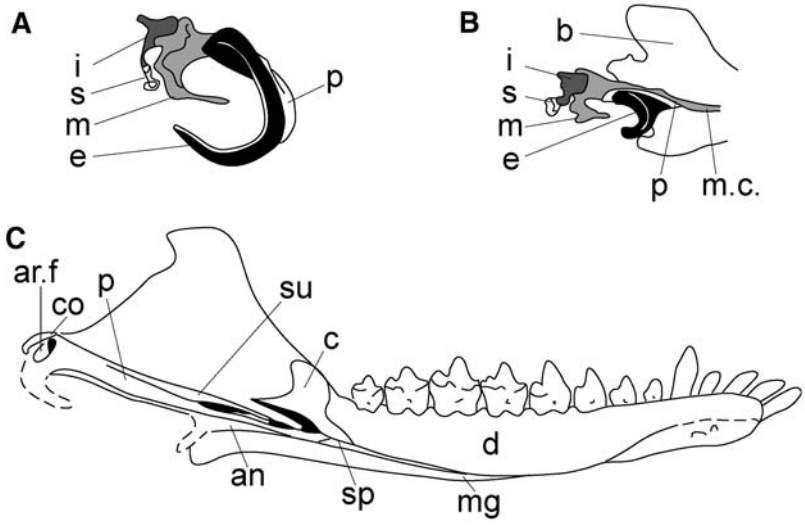


Fig. 1. (**A**) Middle ear ossicles (malleus, incus, and stapes) and tympanic ring (ectotympanic) of an adult opossum *Didelphis marsupialis* in lateral view. (**B**) Medial view of the lower jaw of a pouch-young opossum, showing that the malleus (formed from the fusion of the ossified posterior end of Meckel's cartilage with a dermal bone, the prearticular) and ectotympanic are components of the lower jaw in early development. The middle portion of Meckel's cartilage atrophies in early post-hatching stages of monotremes, post-birth stages of marsupials, and late fetal stages of placentals, severing the connection of the malleus and ectotympanic to the dentary. (**C**) Diagrammatic view of the mandible of the near-mammal (mammaliaform) *Morganucodon*. Present are both the primitive tetrapod jaw joint, which lies between the articular fossa (ar.f) and the quadrate of the upper jaw, and the neomorphic mammalian jaw joint between the dentary condyle (co) and the squamosal bone of the skull. The angular, which bears a hooklike ventral process, the reflected lamina, is homologous with the ectotympanic of mammals; and the articular and prearticular are homologous with the malleus of mammals. The bone covering the meckelian groove is interpreted as a splenial, following (*39*). The surangular, coronoid, and splenial are absent in living mammals. Abbreviations: an, angular; ar.f, articular fossa; c, coronoid; co, dentary condyle; d, dentary; e, ectotympanic; i, incus; m, malleus; m.c., Meckel's cartilage; mg, meckelian groove; p, prearticular; s, stapes; sp, splenial; su, surangular. [(A) and (B) are after (*1*); (C) is redrawn from (*18*)]

[1]Museum Victoria, Post Office Box 666E, Melbourne, Victoria 3001, Australia. [2]School of Geosciences, Post Office Box 28E, Monash University, Victoria 3800, Australia. [3]Department of Organismal Biology and Anatomy, University of Chicago, 1027 East 57th Street, Chicago, IL 60637, USA. [4]Australian Museum, 6 College Street, Sydney, New South Wales 2010, Australia. [5]South Australian Museum, North Terrace, Adelaide, South Australia 5000, Australia.

*To whom correspondence should be addressed. E-mail: trich@museum.vic.gov.au

lage, as it does in living reptiles [(*27*), p. 199]. Because it was unossified, the more anterior portion of Meckel's cartilage is not usually preserved in fossils, but a narrow trough on the dorsal surface of the angular in nonmammalian synapsids indicates its former presence in the living animal [(*18*), figure 35; (*27*), figure 106; (*28*), figures 31 and 32]. In many Mesozoic mammals, a narrow groove, usually called the meckelian groove (*18*), on the medial surface of the dentary below the tooth row is thought to contain the most anterior portion of Meckel's cartilage (*6*, *7*, *18*).

The nature of the postdentary bones in *Teinolophos* cannot be known with certainty, but the morphology of the mandibular trough implies that it housed a rod of accessory jaw bones enclosing an unossified Meckel's cartilage. The lateral wall and roof of the anterior part of the mandibular trough are smoothly curved, forming in cross section the arc of a circle (Fig. 2C). However, the floor of the mandibular trough, from a point about midway along its length forward into the mandibular canal, is a flat surface that meets the curved lateral wall of the trough at a distinct angle (a.f, Fig. 2, C and D). The observation that the floor of the trough is flat and clearly set off from the trough's curved lateral wall suggests that it was a contact surface for an accessory jaw bone. Comparison with *Morganucodon* (*18*) and the docodont *Haldanodon* (*25*, *26*) indicates that this element would be the angular, which contacts the floor of the mandibular trough nearly as far forward as, and somewhat ventral to, the opening of the mandibular foramen. An identifiable facet for the surangular on the lateral and dorsal walls of the trough cannot be distinguished, but such a contact may have existed. The medially located prearticular, however, would not be expected to leave evidence of its former presence on the dentary, except perhaps in the area just anteromedial to the mandibular foramen, which, however, is damaged in this specimen.

The extension of the presumed angular contact of *Teinolophos* into the mandibular canal is not seen in early mammaliaforms but occurs in nonmammalian cynodonts; in the Early Triassic cynodont *Thrinaxodon*, the angular, which bears a dorsal groove for Meckel's cartilage, extends forward within the mandibular canal to the level of the last tooth (*28*). The extension of the angular into the mandibular canal is presumably a secondary condition in *Teinolophos*, possibly associated with the great enlargement of the mandibular foramen.

Evidence for the retention of two additional accessory jaw bones, sometimes referred to

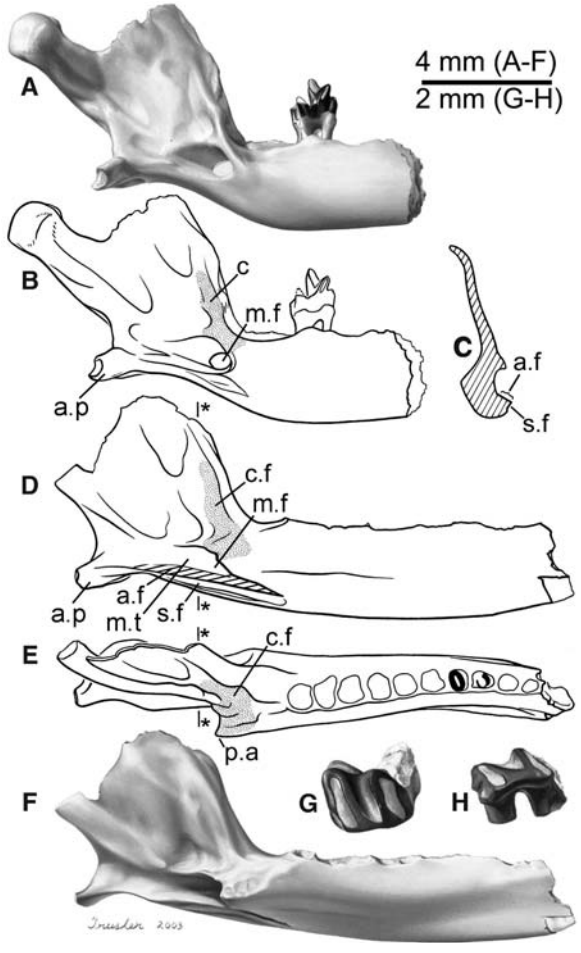


Fig. 2. (**A**) Medial view of holotype of *T. trusleri*, specimen NMV P208231. (**B**) Diagrammatic medial view of NMV P208231; the stippled area indicates the position of the fused coronoid bone. (**C**) Cross section of mandible of referred specimen of *T. trusleri*, NMV P212933; position of cross section is indicated in (D) and (E) by lines terminated with asterisks. (**D**) Diagrammatic medial view of NMV P212933. The stippled area indicates the position of the contact facet for the coronoid bone. Diagonal lines indicate the flat facet interpreted as a contact surface for the angular bone. (**E**) Diagrammatic dorsal view of NMV P212933. Traces of roots of a molar can be seen in alveoli three and four. (**F**) Medial view of NMV P212933, rotated slightly medially toward the viewer. (**G**) Occlusal and (**H**) medial views of isolated lower molar associated with dentary, NMV P212933. The match between the preserved roots of the lower molar found close to alveoli three and four of dentary NMV P212933 and the remnants of roots in those alveoli suggest that is where the tooth was located in the jaw. The presence of a prominent wear facet across the entire width of the tooth on the anterior lophid of the lower molar implies that there was an upper molar anterior to this lower molar and thus the latter cannot be the m1. So the preserved first and second alveoli on the dentary must be for a molar, behind which were four additional molars. Abbreviations: a.f, angular facet; a.p, angular process; c, coronoid; c.f, coronoid facet; m.f, mandibular foramen; m.t, mandibular trough; p.a, posterointernal angle; s.f, splenial facet.

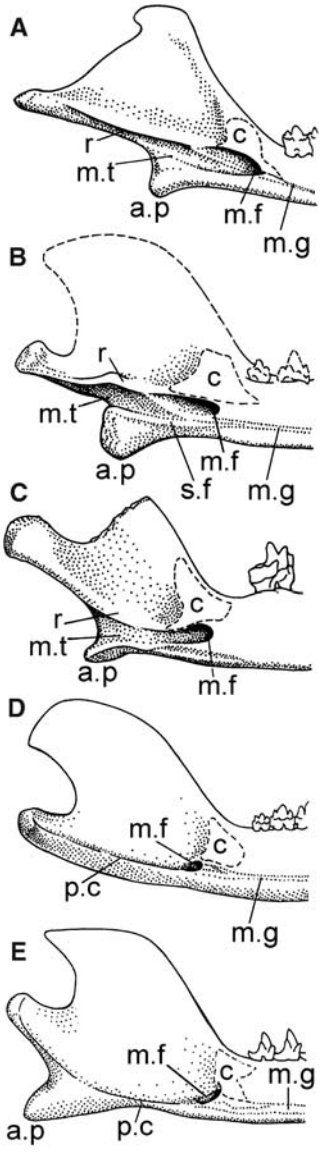


Fig. 3. Comparison of the dentary of *T. trusleri* (**C**) with selected nonmammalian mammaliaformes (**A** and **B**) and mammals (**D** and **E**). (A) *Morganucodon*. (B) *Docodon*. (C) *Teinolophos*. (D) The amphilestid triconodont *Phascolotherium*. (E) The primitive cladothere *Amphitherium*. The angular process of *Amphitherium* (E) and other members of the Cladotheria (*40*), including therians, lies below the condyle rather than well anterior to it and may not be homologous with the angular process of monotremes, basal mammaliaforms, and nonmammalian cynodonts (*14*, *40*, *41*). Abbreviations: a.p, angular process; c, coronoid facet; m.f, mandibular foramen; m.g, meckelian groove; m.t, mandibular trough; p.c, pterygoid crest; r, ridge; s.f, splenial facet. [(A) is after (*18*); (B) and (C) are original; (D) and (E) are after (*30*)]

as paradentary bones (*1*, *6*, *10*), is seen on the most complete referred dentary. On the anteromedial edge of the ascending ramus is a flat facet (c.f, Fig. 2, D to F) in the position of the coronoid bone and its underlying facet in *Morganucodon* (c, Figs. 1C and 3A) and the coronoid facet seen in many early mammaliaforms (c, Fig. 3, B, D, and E). On the holotype and the fragmentary referred specimen NMV P208526, a medially directed flange of bone occupies the same area (c, Fig. 2, A and B). This flange is interpreted here as a coronoid that has fused to the dentary. In the large available sample of *Morganucodon*, some dentaries preserve only the contact facet for the coronoid, whereas others have a bulbous coronoid in place [(*18*), figure 34]. In some cases, the coronoid of *Morganucodon* is fused to the dentary without the trace of a suture (*18*).

In referred specimen NMV P212933 of *T. trusleri*, a shallow bifaceted groove (s.f, Fig. 2, C and D) lies below the anterior part of the mandibular trough, being separated from the trough by a sharp lip, and extends forward nearly to the anterior end of a V-shaped break. The dorsal two-thirds of this groove is formed by a flat surface that faces ventromedially, whereas the lower one-third is formed by a flat medially facing surface. The fragmentary juvenile specimen NMV P212811 preserves a single flat facet in the same position below the mandibular trough. This faceted surface appears to be homologous with a similarly located facet on the dentary that served as a contact for the splenial in Triassic cynodonts, such as *Cynognathus* [(*18*), figures 10 and 26] and, by inference, in the Late Jurassic docodontids, such as *Docodon* (s.f., Fig. 3B) and *Haldanodon* [(*25*), figure 4 and plate 1]. The facet for the presumed splenial in *Teinolophos* extends for a short distance anterior to the mandibular foramen, unlike the usual condition in nonmammalian cynodonts, where it extends to the symphysis, or in *Morganucodon*, where it occupies the entire meckelian groove (*29*). However, *Teinolophos* lacks an anterior extension of the meckelian groove and therefore may have lost the anterior part of the splenial usually associated with this groove. The anterior portion of Meckel's cartilage, considered to lie within the groove (*6*, *7*, *18*), may have been ossified as part of the dentary.

Teinolophos shares with morganucodont and docodont mammaliaforms the following primitive features of the lower jaw (Fig. 3, A to C): a prominent mandibular trough bordered above by a well-developed ridge (implying that the middle ear ossicles and ectotympanic of early monotremes retained an adult contact with the lower jaw); a mandibular foramen lying at the anterior end of the trough below the ridge; and an angular process directed posteroventrally and lying below the lower margin of the jaw well anterior to the articular condyle. The failure of the facet for the angular to extend posteriorly as far as the angular process suggests that the posterior part of the postdentary rod in *Teinolophos* was shifting medially, away from close proximity to the dentary condyle to the more medially located middle ear cavity, a condition to be expected in a transitional form in which the ear bones are beginning to separate from the jaw. In most Mesozoic Mammalia other than monotremes (in triconodonts, multituberculates, symmetrodonts, dryolestoids, and *Amphitherium*), the internal mandibular trough and its overlying ridge are absent and the mandibular foramen opens backward into a broad, flat or slightly concave area bounded ventrally by a prominent ridge (Fig. 3, D and E): the "pterygoid crest" of Simpson (*30*). This ridge extends from the lower margin of the mandibular foramen back either on to the medial side of the angular process (for example, in dryolestoids) or on to the medial side of the condylar process [for example, in triconodonts (Fig. 3D), multituberculates, symmetrodonts, and *Amphitherium* (Fig. 3E)]. The absence of the mandibular trough and its overlying ridge implies that in these taxa the ear ossicles and ectotympanic bone were no longer connected to the lower jaw via a complete Meckel's cartilage. This is known to be the case in living monotremes and therians and in the extinct multituberculates, where definitive ear ossicles are preserved (*5*), and probably in gobiconodontid triconodonts, where an ossified middle portion of Meckel's cartilage appears to be detached from the malleus (*6*, *7*).

The phylogenetic position of monotremes with respect to marsupials and placentals (crown group therians) has long been a point of contention. Both molecular and morphological evidence has been mustered to support either a sister group relationship of monotremes with marsupials (Marsupionta hypothesis) (Fig. 4) or, alternatively, a distant relationship between monotremes and crown group therians [reviewed in (*20*, *31*, *32*)]. The current consensus of molecular and morphological analyses places monotremes outside of a monophyletic Theria (*31*).

As shown in Fig. 4, uncertainty surrounds the relationships of monotremes to other, primarily Mesozoic, mammalian taxa [views summarized in (*20*, *21*)]. It is generally agreed that monotremes are more derived than morganucodontids and docodontids, which are thus excluded from Mammalia and classified as nonmammalian Mammaliaformes (*33*, *34*) (Fig. 4). A number of studies place Monotremata as the outgroup to a clade comprising Triconodonta, Multituberculata, and Trechnotheria (*19*, *34*–*36*), unequivocal members of which lack evidence of middle ear elements attached to the lower jaw (Fig. 3, D and E). This higher-level clade, the Theriimorpha (Fig. 4), is defined by Rowe as a monophyletic taxon that includes therians and all extinct taxa closer to therians than to monotremes (*34*). We provisionally accept this phylogenetic position of monotremes because it is in accord with the polyphyletic origin of the definitive mammalian middle ear but requires the least amount of homoplasy in com-

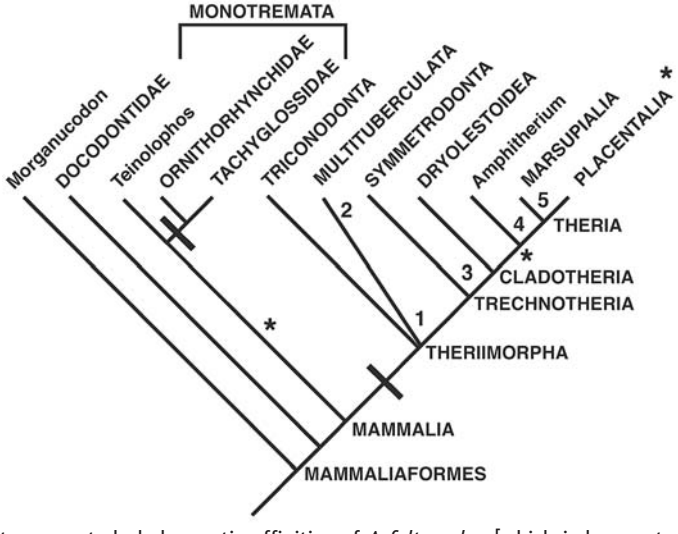


Fig. 4. Cladogram showing the relationship of *Teinolophos* to living monotremes, nonmammalian Mammaliaformes, and theriimorph mammals. Because the relative position of triconodonts and multituberculates with respect to trechnotheres remains controversial (*35*, *42*), we place them in a trichotomy with Trechnotheria. The heavy bars indicate the phylogenetic positions most strongly supported for loss of the mandibular trough and freeing of the mammalian middle ear bones from the dentary. The asterisks indicate suggested phylogenetic affinities of *Asfaltomylos* [which is known to possess a mandibular trough (*16*)] and other Gondwanan taxa that are either clustered as the Australosphenida (*43*) and allied with Monotremata (*16*, *19*); allied with nontribosphenic, presumably trechnotherian, mammals (*44*); or allied with *Erinaceus* within Placentalia (*20*, *21*). Suggested alternative positions of monotremes are: (1) closer to Theria than are Triconodonta (*6*, *20*, *21*, *45*, *46*); (2) sister group of Multituberculata (*3*, *5*, *6*, *46*); (3) near the base of Trechnotheria (*1*, *47*); (4) allied with pretribosphenic cladotheres (*48*); and (5) sister group of Marsupialia (Marsupionta hypothesis) (*49*, *50*). Each of these suggested positions of monotremes requires additional independent origins of the definitive mammalian middle ear, as does the placement of the australosphenids within Trechnotheria (including within Placentalia).

parison with other proposed phylogenetic placements of monotremes (Fig. 4).

As noted earlier, a well-developed mandibular trough, indicative of a complete Meckel's cartilage and postdentary jaw bones contacting the dentary in adult individuals, occurs in a number of Mesozoic mammals (or near-mammals) other than *Teinolophos* (*12*–*16*). However, because of the uncertain phylogenetic positions of these taxa with respect to true mammals (monotremes and theriiforms), none provides unequivocal support for the multiple origin of the definitive mammalian middle ear bones. Nonetheless, they suggest the possibility that the freeing of the mammalian ear bones from the lower jaw may have occurred more often than can be conclusively documented at present. If the postdentary bones were already functioning in hearing in late nonmammalian cynodonts and basal mammaliaforms (*1*, *10*, *37*), then this final step in the functional separation of the mammalian middle ear system from the feeding apparatus would be expected to occur in all later lineages.

References and Notes

1. E. F. Allin, J. A. Hopson, in *The Evolutionary Biology of Hearing*, D. B. Webster, R. R. Fay, A. N. Popper, Eds. (Springer Verlag, New York, 1992), pp. 587–614.
2. C. Patterson, in *Vicariance Biogeography: A Critique*, G. J. Nelson, D. E. Rosen, Eds. (Columbia Univ. Press, New York, 1980), pp. 446–500.
3. T. S. Kemp, *Zool. J. Linn. Soc. London* **77**, 353 (1983).
4. M. J. Novacek, in *The Skull, v. 2, Patterns of Structural and Systematic Diversity*, J. Hanken, Brian K. Hall, Eds. (Univ. of Chicago Press, Chicago, 1993), pp. 438–545.
5. J. Meng, A. R. Wyss, *Nature* **377**, 141 (1995).
6. Y. Wang, Y. Hu, J. Meng, C. Li, *Science* **294**, 357 (2001).
7. J. Meng, Y. Hu, Y. Wang, C. Li, *Zool. J. Linn. Soc. London* **138**, 431 (2003).
8. J. A. Hopson, *Am. Zool.* **6**, 437 (1966).
9. G. Fleischer, *Säugetierk. Mitt.* **21**, 131 (1973).
10. E. F. Allin, *J. Morphol.* **47**, 403 (1975).
11. R. Presley, *Sympos. Zool. Soc. London* **52**, 127 (1984).
12. D. M. Kermack, K. A. Kermack, F. Mussett, *J. Linn. Soc. (Zool.)* **47**, 312 (1968).
13. M. Chow, T. H. Rich, *Aust. Mammal.* **5**, 127 (1982).
14. F. A. Jenkins Jr., A. W. Crompton, W. R. Downs, *Science* **222**, 1233 (1983).
15. F. A. Jenkins Jr., S. M. Gatesy, N. H. Shubin, W. W. Amaral, *Nature* **385**, 715 (1997).
16. O. W. M. Rauhut, T. Martin, E. Ortiz-Jaureguizar, P. Puerta, *Nature* **416**, 165 (2002).
17. C. Reichert, *Müllers Arch. Anat. Physiol. Wiss. Med.* **1837**, 120 (1837).
18. K. A. Kermack, F. Mussett, H. W. Rigney, *Zool. J. Linn. Soc. London* **53**, 87 (1973).
19. Z.-X. Luo, Z. Kielan-Jaworowska, R. L. Cifelli, *Acta Palaeontol. Pol.* **47**, 1 (2002).
20. M. O. Woodburne, T. H. Rich, M. S. Springer, *Mol. Phylog. Evol.* **28**, 360 (2003).
21. M. O. Woodburne, *J. Mamm. Evol.* **10**, 195 (2003).
22. T. Rich *et al.*, *Rec. Queen Vic. Mus.* **106**, 1 (1999).
23. T. Rich *et al.*, *Acta Palaeontol. Pol.* **46**, 113 (2001).
24. The holotype of *T. trusleri* was identified as a monotreme primarily on the basis of its bilophodont lower molar crown pattern, similar to that of the Early Cretaceous (Albian) monotreme *Steropodon* and the Cenozoic platypus *Obdurodon* (*23*). A much worn but recognizably bilophodont molar (Fig. 2, G and H) was found in close association with the referred specimen, NMV P212933, which preserves broken roots in alveoli three and four that are similar in size to those of the isolated tooth. Both the type and the referred specimen also resemble fossil and living platypuses in the unusually large size of the mandibular (inferior alveolar) foramen and canal, and the referred specimen (plus a fragmentary juvenile dentary, NMV P212811) resembles them in preserving a postero-internal angle (p.a, Fig. 2E) that may be homologous with the platypus mylohyoid process (*38*). Despite some proportional differences from the holotype of *T. trusleri*, the most complete referred dentary and the four more fragmentary specimens are provisionally attributed to the same species.
25. G. Krusat, *Mem. Serv. Geol. Portugal* **27**, 1 (1980).
26. J. A. Lillegraven, G. Krusat, *Contr. Geol. Univ. Wyoming Spec. Pap.* **28**, 39 (1991).
27. A. S. Romer, *Osteology of the Reptiles* (Univ. of Chicago Press, Chicago, 1956).
28. S. Fourie, *Ann. S. Afr. Mus.* **65**, 337 (1974).
29. The possibility must be considered that the presumed splenial contact surface of *Teinolophos* may instead be a contact surface for a remnant of Meckel's cartilage, as recently described in gobiconodontid triconodonts (*6*, *7*). The facet on the dentary of gobiconodontids for the ossified Meckel's cartilage is a prominent longitudinal groove below the mandibular foramen that is continuous anteriorly with the slitlike meckelian groove seen in basal mammaliaforms and many early mammals. Although the anterior part of the presumed splenial facet in *Teinolophos* is in a position similar to the depression for the ossified Meckel's cartilage in gobiconodontids, its posterior part, below the mandibular trough, more closely resembles the splenial facet of nonmammalian cynodonts and docodontids, in which the complete Meckel's cartilage would have lain within the mandibular trough, enclosed by the dermal rod of postdentary bones. Furthermore, in the referred juvenile specimen (NMV P212811) of *T. trusleri*, in which one would expect a more prominent groove supporting a more fully developed Meckel's cartilage than in older individuals, the facet below the mandibular trough is flat and not at all excavated. Therefore, we believe that the contact surface below the mandibular trough in *Teinolophos* was for the splenial bone rather than Meckel's cartilage; the latter, as in *Morganucodon* and docodontids, lying more dorsally, within the mandibular trough.
30. G. G. Simpson, *A Catalogue of the Mesozoic Mammalia in the Geological Department of the British Museum* [British Museum (Natural History), London, 1928].
31. J. Phillips, D. Penny, *Mol. Phylogenet. Syst.* **28**, 171 (2003).
32. A. M. Musser, *Comp. Biochem. Phys. A* **136**, 927 (2003).
33. T. Rowe, *J. Vertebr. Paleontol.* **8**, 241 (1988).
34. T. Rowe, in *Mammal Phylogeny, v. 1, Mesozoic Differentiation, Multituberculates, Monotremes, Early Therians, and Marsupials*, F. S. Szalay, M. J. Novacek, M. C. McKenna, Eds. (Springer Verlag, New York, 1993), pp. 129–145.
35. G. W. Rougier, J. R. Wible, J. A. Hopson, *Am. Mus. Novit.* **3183**, 1 (1996).
36. J. A. Hopson, in *Encyclopedia of Paleontology*, R. Singer, Ed. (Fitzroy Dearborn, Chicago, 2000), pp. 691–701.
37. U. Zeller, in *Mammal Phylogeny, v. 1, Mesozoic Differentiation, Multituberculates, Monotremes, Early Therians, and Marsupials*, F. S. Szalay, M. J. Novacek, M. C. McKenna, Eds. (Springer Verlag, New York, 1993), pp. 95–107.
38. A. M. Musser, M. Archer, *Philos. Trans. R. Soc. London Ser. B* **353**, 1063 (1998).
39. F. R. Parrington, *Philos. Trans. R. Soc. London Ser. B* **261**, 231 (1971).
40. B. Patterson, *Fieldiana Geol.* **13**, 1 (1956).
41. D. R. Prothero, *Bull. Am. Mus. Nat. Hist.* **167**, 277 (1981).
42. J. R. Wible, G. W. Rougier, M. J. Novacek, M. C. McKenna, D. D. Dashzeveg, *Am. Mus. Novit.* **3149**, 1 (1995).
43. Z.-X. Luo, R. L. Cifelli, Z. Kielan-Jaworowska, *Nature* **409**, 53 (2001).
44. A. Forasiepi, G. Rougier, A. Martinelli, *J. Vertebr. Paleontol.* **24** (suppl. 3), 59A (2004).
45. Q. Ji, Z.-X. Luo, S. Ji, *Nature* **398**, 326 (1999).
46. J. R. Wible, J. A. Hopson, in *Mammal Phylogeny, v. 1, Mesozoic Differentiation, Multituberculates, Monotremes, Early Therians, and Marsupials*, F. S. Szalay, M. J. Novacek, M. C. McKenna, Eds. (Springer Verlag, New York, 1993), pp. 45–62.
47. J. A. Hopson, in *Major Features of Vertebrate Evolution*, D. R. Prothero, R. M. Schoch, Eds. (Paleontological Society, Knoxville, TN, 1994), pp. 190–219.
48. Z. Kielan-Jaworowska, A. W. Crompton, F. A. Jenkins Jr., *Nature* **326**, 871 (1987).
49. A. Janke, X. Xu, U. Arnason, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 1276 (1997).
50. D. Penny, M. Hasegawa, *Nature* **387**, 549 (1997).

51. Supported jointly by the Committee for Research and Exploration of the National Geographic Society, grant 7370-02, and the Australian Research Council, project DP0209280. The manuscript was critiqued by W. A. Clemens, L. L. Jacobs, M. C. McKenna, J. Meng, P. F. Murray, G. W. Rougier, M. S. Springer, and M. O. Woodburne.

28 September 2004; accepted 22 December 2004
10.1126/science.1105717

Shell Composition Has No Net Impact on Large-Scale Evolutionary Patterns in Mollusks

Susan M. Kidwell

A major suspected bias in the fossil record of skeletonized groups is variation in preservability owing to differences in shell composition. However, despite extensive changes in shell composition over the 500-million-year history of marine bivalves, genus duration and shell composition show few significant relationships, and of those, virtually all are contrary to bias from preferential loss of highly reactive shell types. Distortion of large-scale temporal patterns in marine bivalves owing to preservability is thus apparently weak or randomly distributed, which increases the likelihood that observed patterns in this and other shelled groups carry a strong biological signal.

In the sedimentary record, biological hard parts, inorganic grains, and cements originally composed of calcite are typically better preserved than their aragonitic counterparts (*1*, *2*). This calcite bias, along with experimental evidence that skeletal microstructure can also strongly influence preservation (*3*–*6*), is of concern because of the potential to distort paleoecological patterns among shelled invertebrates, which constitute the bulk of the marine metazoan

Department of Geophysical Sciences, University of Chicago, 5734 South Ellis Avenue, Chicago, IL 60637, USA. E-mail: skidwell@uchicago.edu

parison with other proposed phylogenetic placements of monotremes (Fig. 4).

As noted earlier, a well-developed mandibular trough, indicative of a complete Meckel's cartilage and postdentary jaw bones contacting the dentary in adult individuals, occurs in a number of Mesozoic mammals (or near-mammals) other than *Teinolophos* (*12*–*16*). However, because of the uncertain phylogenetic positions of these taxa with respect to true mammals (monotremes and theriiforms), none provides unequivocal support for the multiple origin of the definitive mammalian middle ear bones. Nonetheless, they suggest the possibility that the freeing of the mammalian ear bones from the lower jaw may have occurred more often than can be conclusively documented at present. If the postdentary bones were already functioning in hearing in late nonmammalian cynodonts and basal mammaliaforms (*1*, *10*, *37*), then this final step in the functional separation of the mammalian middle ear system from the feeding apparatus would be expected to occur in all later lineages.

References and Notes

1. E. F. Allin, J. A. Hopson, in *The Evolutionary Biology of Hearing*, D. B. Webster, R. R. Fay, A. N. Popper, Eds. (Springer Verlag, New York, 1992), pp. 587–614.
2. C. Patterson, in *Vicariance Biogeography: A Critique*, G. J. Nelson, D. E. Rosen, Eds. (Columbia Univ. Press, New York, 1980), pp. 446–500.
3. T. S. Kemp, *Zool. J. Linn. Soc. London* **77**, 353 (1983).
4. M. J. Novacek, in *The Skull, v. 2, Patterns of Structural and Systematic Diversity*, J. Hanken, Brian K. Hall, Eds. (Univ. of Chicago Press, Chicago, 1993), pp. 438–545.
5. J. Meng, A. R. Wyss, *Nature* **377**, 141 (1995).
6. Y. Wang, Y. Hu, J. Meng, C. Li, *Science* **294**, 357 (2001).
7. J. Meng, Y. Hu, Y. Wang, C. Li, *Zool. J. Linn. Soc. London* **138**, 431 (2003).
8. J. A. Hopson, *Am. Zool.* **6**, 437 (1966).
9. G. Fleischer, *Säugetierk. Mitt.* **21**, 131 (1973).
10. E. F. Allin, *J. Morphol.* **47**, 403 (1975).
11. R. Presley, *Sympos. Zool. Soc. London* **52**, 127 (1984).
12. D. M. Kermack, K. A. Kermack, F. Mussett, *J. Linn. Soc. (Zool.)* **47**, 312 (1968).
13. M. Chow, T. H. Rich, *Aust. Mammal.* **5**, 127 (1982).
14. F. A. Jenkins Jr., A. W. Crompton, W. R. Downs, *Science* **222**, 1233 (1983).
15. F. A. Jenkins Jr., S. M. Gatesy, N. H. Shubin, W. W. Amaral, *Nature* **385**, 715 (1997).
16. O. W. M. Rauhut, T. Martin, E. Ortiz-Jaureguizar, P. Puerta, *Nature* **416**, 165 (2002).
17. C. Reichert, *Müllers Arch. Anat. Physiol. Wiss. Med.* **1837**, 120 (1837).
18. K. A. Kermack, F. Mussett, H. W. Rigney, *Zool. J. Linn. Soc. London* **53**, 87 (1973).
19. Z.-X. Luo, Z. Kielan-Jaworowska, R. L. Cifelli, *Acta Palaeontol. Pol.* **47**, 1 (2002).
20. M. O. Woodburne, T. H. Rich, M. S. Springer, *Mol. Phylog. Evol.* **28**, 360 (2003).
21. M. O. Woodburne, *J. Mamm. Evol.* **10**, 195 (2003).
22. T. Rich *et al.*, *Rec. Queen Vic. Mus.* **106**, 1 (1999).
23. T. Rich *et al.*, *Acta Palaeontol. Pol.* **46**, 113 (2001).
24. The holotype of *T. trusleri* was identified as a monotreme primarily on the basis of its bilophodont lower molar crown pattern, similar to that of the Early Cretaceous (Albian) monotreme *Steropodon* and the Cenozoic platypus *Obdurodon* (*23*). A much worn but recognizably bilophodont molar (Fig. 2, G and H) was found in close association with the referred specimen, NMV P212933, which preserves broken roots in alveoli three and four that are similar in size to those of the isolated tooth. Both the type and the referred specimen also resemble fossil and living platypuses in the unusually large size of the mandibular (inferior alveolar) foramen and canal, and the referred specimen (plus a fragmentary juvenile dentary, NMV P212811) resembles them in preserving a postero-internal angle (p.a, Fig. 2E) that may be homologous with the platypus mylohyoid process (*38*). Despite some proportional differences from the holotype of *T. trusleri*, the most complete referred dentary and the four more fragmentary specimens are provisionally attributed to the same species.
25. G. Krusat, *Mem. Serv. Geol. Portugal* **27**, 1 (1980).
26. J. A. Lillegraven, G. Krusat, *Contr. Geol. Univ. Wyoming Spec. Pap.* **28**, 39 (1991).
27. A. S. Romer, *Osteology of the Reptiles* (Univ. of Chicago Press, Chicago, 1956).
28. S. Fourie, *Ann. S. Afr. Mus.* **65**, 337 (1974).
29. The possibility must be considered that the presumed splenial contact surface of *Teinolophos* may instead be a contact surface for a remnant of Meckel's cartilage, as recently described in gobiconodontid triconodonts (*6*, *7*). The facet on the dentary of gobiconodontids for the ossified Meckel's cartilage is a prominent longitudinal groove below the mandibular foramen that is continuous anteriorly with the slitlike meckelian groove seen in basal mammaliaforms and many early mammals. Although the anterior part of the presumed splenial facet in *Teinolophos* is in a position similar to the depression for the ossified Meckel's cartilage in gobiconodontids, its posterior part, below the mandibular trough, more closely resembles the splenial facet of nonmammalian cynodonts and docodontids, in which the complete Meckel's cartilage would have lain within the mandibular trough, enclosed by the dermal rod of postdentary bones. Furthermore, in the referred juvenile specimen (NMV P212811) of *T. trusleri*, in which one would expect a more prominent groove supporting a more fully developed Meckel's cartilage than in older individuals, the facet below the mandibular trough is flat and not at all excavated. Therefore, we believe that the contact surface below the mandibular trough in *Teinolophos* was for the splenial bone rather than Meckel's cartilage; the latter, as in *Morganucodon* and docodontids, lying more dorsally, within the mandibular trough.
30. G. G. Simpson, *A Catalogue of the Mesozoic Mammalia in the Geological Department of the British Museum* [British Museum (Natural History), London, 1928].
31. J. Phillips, D. Penny, *Mol. Phylogenet. Syst.* **28**, 171 (2003).
32. A. M. Musser, *Comp. Biochem. Phys. A* **136**, 927 (2003).
33. T. Rowe, *J. Vertebr. Paleontol.* **8**, 241 (1988).
34. T. Rowe, in *Mammal Phylogeny, v. 1, Mesozoic Differentiation, Multituberculates, Monotremes, Early Therians, and Marsupials*, F. S. Szalay, M. J. Novacek, M. C. McKenna, Eds. (Springer Verlag, New York, 1993), pp. 129–145.
35. G. W. Rougier, J. R. Wible, J. A. Hopson, *Am. Mus. Novit.* **3183**, 1 (1996).
36. J. A. Hopson, in *Encyclopedia of Paleontology*, R. Singer, Ed. (Fitzroy Dearborn, Chicago, 2000), pp. 691–701.
37. U. Zeller, in *Mammal Phylogeny, v. 1, Mesozoic Differentiation, Multituberculates, Monotremes, Early Therians, and Marsupials*, F. S. Szalay, M. J. Novacek, M. C. McKenna, Eds. (Springer Verlag, New York, 1993), pp. 95–107.
38. A. M. Musser, M. Archer, *Philos. Trans. R. Soc. London Ser. B* **353**, 1063 (1998).
39. F. R. Parrington, *Philos. Trans. R. Soc. London Ser. B* **261**, 231 (1971).
40. B. Patterson, *Fieldiana Geol.* **13**, 1 (1956).
41. D. R. Prothero, *Bull. Am. Mus. Nat. Hist.* **167**, 277 (1981).
42. J. R. Wible, G. W. Rougier, M. J. Novacek, M. C. McKenna, D. D. Dashzeveg, *Am. Mus. Novit.* **3149**, 1 (1995).
43. Z.-X. Luo, R. L. Cifelli, Z. Kielan-Jaworowska, *Nature* **409**, 53 (2001).
44. A. Forasiepi, G. Rougier, A. Martinelli, *J. Vertebr. Paleontol.* **24** (suppl. 3), 59A (2004).
45. Q. Ji, Z.-X. Luo, S. Ji, *Nature* **398**, 326 (1999).
46. J. R. Wible, J. A. Hopson, in *Mammal Phylogeny, v. 1, Mesozoic Differentiation, Multituberculates, Monotremes, Early Therians, and Marsupials*, F. S. Szalay, M. J. Novacek, M. C. McKenna, Eds. (Springer Verlag, New York, 1993), pp. 45–62.
47. J. A. Hopson, in *Major Features of Vertebrate Evolution*, D. R. Prothero, R. M. Schoch, Eds. (Paleontological Society, Knoxville, TN, 1994), pp. 190–219.
48. Z. Kielan-Jaworowska, A. W. Crompton, F. A. Jenkins Jr., *Nature* **326**, 871 (1987).
49. A. Janke, X. Xu, U. Arnason, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 1276 (1997).
50. D. Penny, M. Hasegawa, *Nature* **387**, 549 (1997).

51. Supported jointly by the Committee for Research and Exploration of the National Geographic Society, grant 7370-02, and the Australian Research Council, project DP0209280. The manuscript was critiqued by W. A. Clemens, L. L. Jacobs, M. C. McKenna, J. Meng, P. F. Murray, G. W. Rougier, M. S. Springer, and M. O. Woodburne.

28 September 2004; accepted 22 December 2004
10.1126/science.1105717

Shell Composition Has No Net Impact on Large-Scale Evolutionary Patterns in Mollusks

Susan M. Kidwell

A major suspected bias in the fossil record of skeletonized groups is variation in preservability owing to differences in shell composition. However, despite extensive changes in shell composition over the 500-million-year history of marine bivalves, genus duration and shell composition show few significant relationships, and of those, virtually all are contrary to bias from preferential loss of highly reactive shell types. Distortion of large-scale temporal patterns in marine bivalves owing to preservability is thus apparently weak or randomly distributed, which increases the likelihood that observed patterns in this and other shelled groups carry a strong biological signal.

In the sedimentary record, biological hard parts, inorganic grains, and cements originally composed of calcite are typically better preserved than their aragonitic counterparts (*1*, *2*). This calcite bias, along with experimental evidence that skeletal microstructure can also strongly influence preservation (*3*–*6*), is of concern because of the potential to distort paleoecological patterns among shelled invertebrates, which constitute the bulk of the marine metazoan

Department of Geophysical Sciences, University of Chicago, 5734 South Ellis Avenue, Chicago, IL 60637, USA. E-mail: skidwell@uchicago.edu

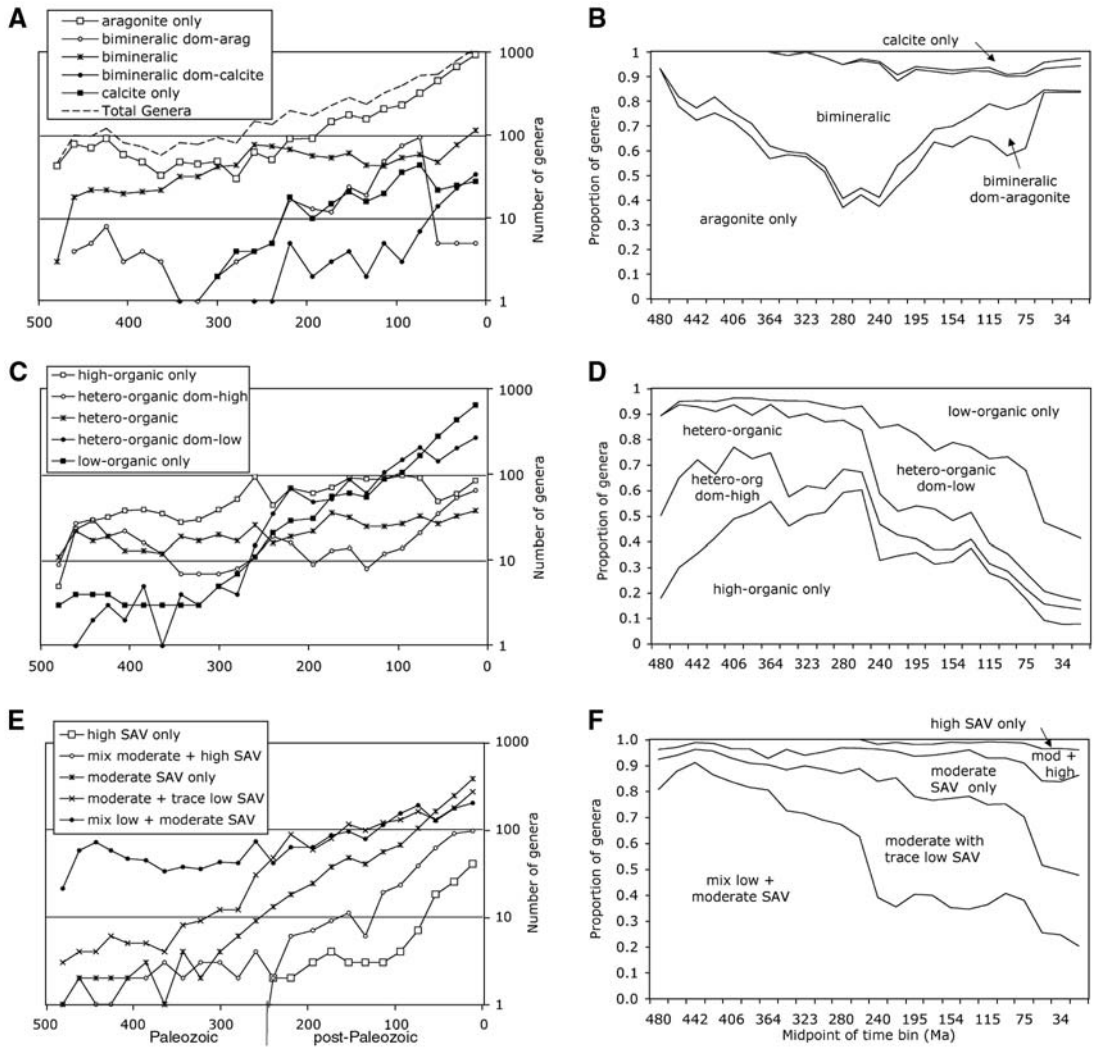


Fig. 1. Trends over Phanerozoic time in the raw number and proportions of marine bivalve genera categorized by shell mineralogy (**A** and **B**), organic content (**C** and **D**), and SAV of first-order crystallites (**E** and **F**), as shown by 20-million-year binning of genus range data.

fossil record (*7*–*10*). Differential preservation of taxa is also a likely source of bias in large-scale evolutionary patterns and is expected to truncate observed geologic durations, damp diversity patterns, and compromise taxonomic identifications (*6*, *11*–*17*). However, its quantitative impact on large-scale evolutionary patterns remains largely unexplored (*6*, *18*).

Preservational biases are evaluated here with the use of bivalve mollusks, an excellent test case because of their compositionally diverse shells. I test (i) whether observed changes in shell composition through the Phanerozoic are consistent with cumulative loss of highly reactive shell types, (ii) whether singleton taxa [confined to a single chronostratigraphic interval and often attributed to sampling deficiencies; reviewed in (*19*)] tend to have more reactive shells than nonsingleton taxa, and (iii) whether median durations of genera having highly reactive shells are shorter than those of genera with less reactive shells.

Information on shell composition compiled from the literature (for 475 genera, 156 families or subfamilies) was quantified by scoring the shell of each genus on a set of 3-point scales for its overall mineralogy, organic content, and crystallite surface area to volume ratio (SAV) (*6*) (figs. S1 and S2; tables S1 and S2); the results were merged with an existing database of the known stratigraphic ranges of all marine bivalve genera having fossil records [(*20*), as updated by (*21*)]. Because shell composition tends to be highly conserved both among species within a genus and among genera within a family (*22*), it is possible to extrapolate from these 475 genera to 89% of the 2983 genera in the stratigraphic-range database, using published familial assignments and using a consensus method when confamilial genera have conflicting shell compositions (*6*).

A literal reading of the fossil record indicates that genus-level diversity for the class Bivalvia (dashed line in Fig. 1A) is fairly steady from its initial diversification until the late Paleozoic, when it begins a steep rise that continues through the post-Paleozoic to a present-day diversity of ~1300 living genera, of which ~1000 have a known fossil record (*21*, *23*). If diversity trends were determined largely by the intrinsic preservability of shell types—which is a worst-case scenario—then, working backward in time from the present day, groups with low preservation potential (shells composed entirely or largely of aragonite, high-organic, or high surface area microstructures) should be missing more of their early members than groups with high preservation potential. This would steepen their diversity trajectories, diminishing their proportional diversity when tracked back in time, and, if the bias were sufficiently severe, shifting their first occurrences to younger dates relative to groups with high preservation potential.

Instead, observed changes in the generic diversity and proportional representation of preservational groups are largely inconsistent with broad-scale preservational bias. Preservational groups exhibit a variety of discordant and even anastomosing trajectories in raw diversity (Fig. 1, A, C, and E). Moreover, proportional data (which are less subject to bias from temporal variation in sampling intensity; Fig. 1, B, D, and F) show that compositional differences between younger and older faunas are largely contrary to taphonomic expectations. For example, the early and mid-

Paleozoic are dominated by genera with entirely aragonitic shells and with exclusively to dominantly high-organic microstructures (Fig. 1, B and D). The presumably most durable calcite-bearing taxa have their highest proportional richness in the late Paleozoic and early Mesozoic, whereas genera with low-organic shells only rise to dominance in Cenozoic rocks. In contrast, trends in crystallite SAVs are consistent with preservational bias, although they could also reflect phylogeny. Early Paleozoic bivalve faunas are dominated by genera with significant proportions of low-SAV crystallites in their shells, whereas the Mesozoic [250 to 65 million years ago (Ma)] is dominated by shells composed largely or entirely of moderate-SAV microstructures, and the Cenozoic (since 65 Ma) sees a continued increase in the proportions of moderate- and high-SAV microstructures (Fig. 1F).

A complex ad hoc history of changing environmental and diagenetic conditions would be required to generate this set of patterns artifactually: Changes in bivalve shell mineralogy do not match postulated alternations of aragonite and calcite seawaters [as noted by (*24*)], and, although trends in the proportional diversity of high-organic genera are in crude agreement with numbers of silicified benthic faunas [numbers from (*25*)], no other shell characters nor any raw diversity trends agree. Trends are thus more likely biological in origin, with shell composition evolving under selection to provide biomechanical advantage against increasing durophagous predation (*26*, *27*) and/or as an aspect of increasing metabolic efficiency (*28*), given that organic matrix is more costly to deposit than the mineral phase (*29*). Whatever the drivers of the trends, Fig. 1 reveals that the Permian-Triassic boundary (250 Ma) was a watershed in shell composition: Through the Paleozoic (>250 Ma), bivalves show declining proportions of entirely aragonitic taxa, increasing proportions of high-organic shells, and high proportions of shells containing relatively low-SAV microstructures; in the post-Paleozoic (<250 Ma), bivalves show increasing proportions of entirely aragonitic taxa, decreasing proportions of high-organic shells, and a relatively even mix of low-, moderate-, and high-SAV microstructures (Fig. 1, B, D, and F). Given these changes, the presence of only 42 genera shared between Paleozoic and post-Paleozoic intervals, and the much higher diversity of post-Paleozoic bivalves (2471 genera versus 554 in the Paleozoic), the database is partitioned into two separate 250-million-year intervals in tests for preservational bias using singleton status and genus durations.

If a highly reactive shell decreases genus-level preservation potential, and if singleton status derives from poor preservation, then singletons should exhibit greater frequencies of highly reactive shells than nonsingleton taxa. However, the frequency of highly reactive shells among singleton genera differs significantly from their frequency among nonsingleton genera for only five of the nine comparisons (Table 1) (fig. S3), and of these five, only two are consistent with taphonomic expectations (singletons have proportionally more high-organic shells than do nonsingletons, both within the post-Paleozoic subset and in the Phanerozoic overall). Three of the significant differences are opposite to expectations: Singletons have proportionally fewer entirely aragonitic shells in the post-Paleozoic and Phanerozoic and fewer dominantly high-SAV shells in the Phanerozoic than do nonsingletons. Rearranging the question to ask "How similar are preservational groups in their proportions of singletons?" yields a similar pattern, with either no significant difference (five tests) or significant differences that, in three of the remaining four tests, are contrary to preservational bias (table S3 and fig. S4).

If a highly reactive shell decreases genus-level preservation potential, then genera having highly reactive shells should have shorter median durations than genera with other shell compositions. However, regardless of (i) the inclusion or exclusion of singleton taxa, (ii) whether genera are considered in toto or partitioned into Paleozoic and post-Paleozoic subsets, or (iii) whether one considers individual compositional aspects of shells or combined scores (24 nested comparisons, of which 20 compare medians for >100 genera and are analyzed here), median stratigraphic durations of genera either do not differ significantly as a function of shell composition or differ significantly but opposite to taphonomic expectations (fall below the diagonal in Fig. 2) (table S4). Of 18 nonsignificant comparisons, 15 also have differences that are opposite to taphonomic expectations.

Finer partitioning of the database, so that mineralogy and organic content can be considered simultaneously (those aspects most reliably estimated from the literature; lack of multivariate normality disallows multiple regression), reinforces the results of the preceding single-factor analyses (table S5). Variation among post-Paleozoic genera is opposite to taphonomic expectation: Genera with resistant shells (i.e., containing some quantity of calcite and composed entirely of low-organic microstructures) have a significantly shorter median duration and higher proportion of singletons than do genera with shells composed exclusively of high-organic aragonite. The implication is that first-order patterns in post-Paleozoic genus ranges are robust to preservational bias imposed by variation in shell composition; these taxa constitute ~85% of Phanerozoic bivalve diversity. In contrast, variation among Paleozoic

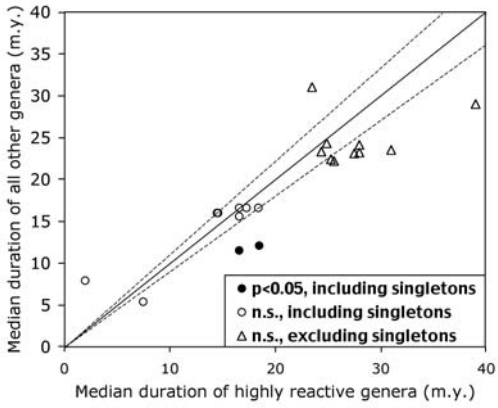


Fig. 2. Bivalves having highly reactive shell compositions—that is, composed entirely of aragonite or high-organic microstructures, dominated by high-SAV microstructures, or having shells that combine these attributes—tend to have median genus durations that are longer than other genera (fall below the diagonal line), contrary to the taphonomic expectation that their durations would be significantly shorter. Dashed lines indicate 10% divergence from equivalence in duration.

Table 1. Results of *G* tests for differences in the proportions of singletons and nonsingletons that have highly reactive shell compositions [see (*6*) for operational definitions]. Min, mineralogy; Org, organic content; SAV, surface area:volume ratio; Phan, Phanerozoic; Pz, Paleozoic; post-Pz, post-Paleozoic.

Aspect tested	Time bin	Singleton genera		Nonsingleton genera		Observed *P* value	Sequential Bonferroni (corrected) *P* value	Difference consistent with taphonomic expectation?
		N	% with highly reactive shell	*N*	% with highly reactive shell			
Min	Phan	671	61	2175	73	0.000001	0.000008	No, opposite
Min	Post-Pz	454	58	1952	74	0.000001	0.000007	No, opposite
Org	Post-Pz	454	17	1943	14	0.000001	0.000006	Yes
Org	Phan	646	27	2146	18	0.0001	0.0005	Yes
SAV	Phan	605	6	2056	10	0.002	0.008	No, opposite
Min	Pz	217	67	262	56	0.018	n.s. (0.054)	Yes
SAV	Post-Pz	437	8	1881	11	0.055	n.s.	No, opposite
Org	Pz	192	50	240	52	0.7	n.s.	No, opposite
SAV	Pz	168	2	210	2	0.95	n.s.	No

genera is consistent with preservational bias—low-organic calcitic taxa have a longer median duration and a smaller proportion of singletons—but differences are not statistically significant (table S5). The magnitude of bias among Paleozoic taxa is thus evidently small [single-factor tests indicate that mineralogy is the strongest factor but is always nonsignificant in impact (tables S3 and S4)].

This analysis does not eliminate potential bias generated by other intrinsic (e.g., body size, original abundance, life habit) or extrinsic factors in preservation [e.g., available rock volume, degree of lithification (*11*–*17*)]. However, these results indicate that bias in large-scale evolutionary patterns related to shell composition is weak at best among marine bivalves or is so complex as to be effectively random. That is, although all observed durations and diversities may be offset downward from their true (original, biological) values—perhaps more severely in the Paleozoic, given the direction of the nonsignificant difference in durations—there appears to be no systematic distortion in those values among genera as a function of shell composition. Differences in durations among bivalve clades are thus more likely to be biologically real than artifacts of differences in shell composition, even though shell composition tends to follow phylogenetic lines. These results are very encouraging for biological analysis of the fossil record at this scale, including the retention of singleton data in evolutionary analyses. Preservational distortion within this major group is not as large as might be feared from short-term experiments (*3*–*5*) or from the striking diversity contrasts that can occur between adjacent beds (*7*–*10*). These preservational differences, however real at the level of individual assemblages, apparently do not scale upward to macroevolutionary patterns.

The apparent lack of bias in genus duration data may have several explanations, which are not mutually exclusive. First, there may be considerable redundancy in preservation (*10*): Although many beds may lack aragonitic taxa entirely, random beds with adequate preservation (composite molds and mineral replacements with high morphologic acuity) might be sufficiently densely spaced for the geologic ranges of taxa with reactive shells to be recognized as accurately as those of less reactive taxa. Second, highly reactive shells might be especially prone to replacement by more stable minerals, actually increasing rather than decreasing their per capita preservation potential [references in (*4*, *30*)]. Third, highly reactive shells commonly have poor preservational quality even when diagenetically replicated (coarse crystalline textures, extraneous coatings) and thus might have lower taxonomic acuity and tend to be assigned to existing named taxa ("lumping"). This would promote artifactual range extension and thereby reduce rather than increase observed percentages of singletons among taxa with highly reactive shells, contrary to the usual algorithm for modeling preservational effects (*6*).

Many skeletonized macroinvertebrate groups in the fossil record have compositions lying within the range spanned by bivalves, and thus there is reason to suspect that differences in skeletal mineralogy and microstructure might have as little net impact on first-order evolutionary patterns as found here among marine bivalves.

References and Notes

1. D. Sanders, *J. Afr. Earth Sci.* **36**, 99 (2003).
2. J. W. Morse, A. Mucci, F. J. Millero, *Geochim. Cosmochim. Acta* **44**, 85 (1980).
3. L. M. Walter, *SEPM Spec. Publ. 36*, 3 (1985).
4. C. P. Glover, S. M. Kidwell, *J. Geol.* **101**, 729 (1993).
5. E. M. Harper, *J. Zool.* **251**, 179 (2000).
6. See supporting data on *Science* Online.
7. C. F. Koch, N. F. Sohl, *Paleobiology* **9**, 26 (1983).
8. L. Cherns, V. P. Wright, *Geology* **28**, 791 (2000).
9. V. P. Wright, L. Cherns, P. Hodges, *Geology* **31**, 211 (2003).
10. A. M. Bush, R. K. Bambach, *J. Geol.* **112**, 625 (2004).
11. D. M. Raup, *Bull. Carnegie Mus. Nat. Hist.* **13**, 85 (1979).
12. M. Kowalewski, K. W. Flessa, *Geology* **24**, 977 (1996).
13. S. K. Donovan, C. R. C. Paul, Eds., *The Adequacy of the Fossil Record* (Wiley, Chichester, UK, 1998).
14. A. B. Smith, *Philos. Trans. R. Soc. London Ser. B* **356**, 351 (2001).
15. M. Foote, in *Paleobiology II, a Synthesis*, D. E. G. Briggs, P. R. Crowther, Eds. (Blackwell, Oxford, 2001), pp. 500–504.
16. A. B. Smith, in *Paleobiology II, a Synthesis*, D. E. G. Briggs, P. R. Crowther, Eds. (Blackwell, Oxford, 2001), pp. 504–509.
17. S. M. Kidwell, S. M. Holland, *Annu. Rev. Ecol. Syst.* **33**, 561 (2002).
18. E. M. Harper, in *The Adequacy of the Fossil Record*, S. K. Donovan, C. R. C. Paul, Eds. (Wiley, Chichester, UK, 1998), pp. 243–267.
19. M. E. Foote, *Paleobiology* **26**, 578 (2000).
20. J. J. Sepkoski Jr., *Bull. Am. Paleontol.* **363**, 1 (2002).
21. D. Jablonski, K. Roy, J. W. Valentine, R. M. Price, P. S. Anderson, *Science* **300**, 1133 (2003).
22. J. D. Taylor, W. J. Kennedy, A. Hall, *Bull. Br. Mus. (Nat. Hist.) Zool. Suppl.* **3**, 1 (1969).
23. A. I. Miller, J. J. Sepkoski Jr., *Paleobiology* **14**, 364 (1998).
24. S. M. Stanley, L. A. Hardie, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **144**, 3 (1998).
25. J. K. Schubert, D. L. Kidder, D. H. Erwin, *Geology* **25**, 1031 (1997).
26. S. A. Wainwright, *Nature* **224**, 777 (1969).
27. M. Kowalewski, P. H. Kelley, Eds., *The Fossil Record of Predation* (Paleontological Society Papers *8*) (Paleontological Society, Lancaster, PA, 2002).
28. R. K. Bambach, *Paleobiology* **19**, 372 (1993).
29. A. R. Palmer, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1379 (1992).
30. R. G. Maliva, R. Siever, *J. Geol.* **96**, 387 (1988).

31. I thank T. R. Waller, D. Jablonski, and J. Pojeta for stimulating discussions and D. Jablonski, M. Foote, M. Labarbera, T. A. Rothfus, M. Kowalewski, and two anonymous persons for helpful reviews. Analyses used the bivalve stratigraphic range data of J. J. Sepkoski Jr. as updated by D. Jablonski (version June 2004). Supported by NSF-EAR grant 0345897.

Supporting Online Material

www.sciencemag.org/cgi/content/full/307/5711/914/DC1
Materials and Methods
Figs. S1 to S4
Tables S1 to S5

22 October 2004; accepted 21 December 2004
10.1126/science.1106654

Two Abundant Bioaccumulated Halogenated Compounds Are Natural Products

Emma L. Teuten, Li Xu, Christopher M. Reddy*

Methoxylated polybrominated diphenyl ethers (MeO-PBDEs) have been found bioaccumulated in the tissues of a variety of aquatic animals and at concentrations comparable to those of anthropogenic halogenated organic compounds, including polychlorinated biphenyls (PCBs). The origin of the MeO-PBDEs has been uncertain; circumstantial evidence supports a natural and/or an industrial source. By analyzing the natural abundance radiocarbon content of two MeO-PBDEs isolated from a True's beaked whale (*Mesoplodon mirus*), we show that these compounds were naturally produced.

It has long been known that industrially produced halogenated organic compounds and their metabolites accumulate in human and other animal tissues. One example is polybrominated diphenyl ethers (PBDEs), used as flame retardants, which have emerged as a contaminant of concern because of the detection of increasing levels in human breast milk (*1*). Structurally related compounds, methoxylated polybrominated diphenyl ethers (MeO-PBDEs), have recently been identified in a variety of fish (*2*) and marine mammals (*3*); they form a new class of bioaccumulated compounds. The origin of these MeO-PBDEs is not immediately obvious. Many MeO-PBDEs are known natural products, isolated from sponges (*4*–*9*), algae (*10*), and acorn worms (*11*). However, their structural similarity to industrial PBDEs cannot be ignored. Industrially produced PBDEs lack a methoxy group, but this group could be introduced by

Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA.

*To whom correspondence should be addressed. E-mail: creddy@whoi.edu

genera is consistent with preservational bias—low-organic calcitic taxa have a longer median duration and a smaller proportion of singletons—but differences are not statistically significant (table S5). The magnitude of bias among Paleozoic taxa is thus evidently small [single-factor tests indicate that mineralogy is the strongest factor but is always nonsignificant in impact (tables S3 and S4)].

This analysis does not eliminate potential bias generated by other intrinsic (e.g., body size, original abundance, life habit) or extrinsic factors in preservation [e.g., available rock volume, degree of lithification (*11*–*17*)]. However, these results indicate that bias in large-scale evolutionary patterns related to shell composition is weak at best among marine bivalves or is so complex as to be effectively random. That is, although all observed durations and diversities may be offset downward from their true (original, biological) values—perhaps more severely in the Paleozoic, given the direction of the nonsignificant difference in durations—there appears to be no systematic distortion in those values among genera as a function of shell composition. Differences in durations among bivalve clades are thus more likely to be biologically real than artifacts of differences in shell composition, even though shell composition tends to follow phylogenetic lines. These results are very encouraging for biological analysis of the fossil record at this scale, including the retention of singleton data in evolutionary analyses. Preservational distortion within this major group is not as large as might be feared from short-term experiments (*3*–*5*) or from the striking diversity contrasts that can occur between adjacent beds (*7*–*10*). These preservational differences, however real at the level of individual assemblages, apparently do not scale upward to macroevolutionary patterns.

The apparent lack of bias in genus duration data may have several explanations, which are not mutually exclusive. First, there may be considerable redundancy in preservation (*10*): Although many beds may lack aragonitic taxa entirely, random beds with adequate preservation (composite molds and mineral replacements with high morphologic acuity) might be sufficiently densely spaced for the geologic ranges of taxa with reactive shells to be recognized as accurately as those of less reactive taxa. Second, highly reactive shells might be especially prone to replacement by more stable minerals, actually increasing rather than decreasing their per capita preservation potential [references in (*4*, *30*)]. Third, highly reactive shells commonly have poor preservational quality even when diagenetically replicated (coarse crystalline textures, extraneous coatings) and thus might have lower taxonomic acuity and tend to be assigned to existing named taxa ("lumping"). This would promote artifactual range extension and thereby reduce rather than increase observed percentages of singletons among taxa with highly reactive shells, contrary to the usual algorithm for modeling preservational effects (*6*).

Many skeletonized macroinvertebrate groups in the fossil record have compositions lying within the range spanned by bivalves, and thus there is reason to suspect that differences in skeletal mineralogy and microstructure might have as little net impact on first-order evolutionary patterns as found here among marine bivalves.

References and Notes

1. D. Sanders, *J. Afr. Earth Sci.* **36**, 99 (2003).
2. J. W. Morse, A. Mucci, F. J. Millero, *Geochim. Cosmochim. Acta* **44**, 85 (1980).
3. L. M. Walter, *SEPM Spec. Publ. 36*, 3 (1985).
4. C. P. Glover, S. M. Kidwell, *J. Geol.* **101**, 729 (1993).
5. E. M. Harper, *J. Zool.* **251**, 179 (2000).
6. See supporting data on *Science* Online.
7. C. F. Koch, N. F. Sohl, *Paleobiology* **9**, 26 (1983).
8. L. Cherns, V. P. Wright, *Geology* **28**, 791 (2000).
9. V. P. Wright, L. Cherns, P. Hodges, *Geology* **31**, 211 (2003).
10. A. M. Bush, R. K. Bambach, *J. Geol.* **112**, 625 (2004).
11. D. M. Raup, *Bull. Carnegie Mus. Nat. Hist.* **13**, 85 (1979).
12. M. Kowalewski, K. W. Flessa, *Geology* **24**, 977 (1996).
13. S. K. Donovan, C. R. C. Paul, Eds., *The Adequacy of the Fossil Record* (Wiley, Chichester, UK, 1998).
14. A. B. Smith, *Philos. Trans. R. Soc. London Ser. B* **356**, 351 (2001).
15. M. Foote, in *Paleobiology II, a Synthesis*, D. E. G. Briggs, P. R. Crowther, Eds. (Blackwell, Oxford, 2001), pp. 500–504.
16. A. B. Smith, in *Paleobiology II, a Synthesis*, D. E. G. Briggs, P. R. Crowther, Eds. (Blackwell, Oxford, 2001), pp. 504–509.
17. S. M. Kidwell, S. M. Holland, *Annu. Rev. Ecol. Syst.* **33**, 561 (2002).
18. E. M. Harper, in *The Adequacy of the Fossil Record*, S. K. Donovan, C. R. C. Paul, Eds. (Wiley, Chichester, UK, 1998), pp. 243–267.
19. M. E. Foote, *Paleobiology* **26**, 578 (2000).
20. J. J. Sepkoski Jr., *Bull. Am. Paleontol.* **363**, 1 (2002).
21. D. Jablonski, K. Roy, J. W. Valentine, R. M. Price, P. S. Anderson, *Science* **300**, 1133 (2003).
22. J. D. Taylor, W. J. Kennedy, A. Hall, *Bull. Br. Mus. (Nat. Hist.) Zool. Suppl.* **3**, 1 (1969).
23. A. I. Miller, J. J. Sepkoski Jr., *Paleobiology* **14**, 364 (1998).
24. S. M. Stanley, L. A. Hardie, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **144**, 3 (1998).
25. J. K. Schubert, D. L. Kidder, D. H. Erwin, *Geology* **25**, 1031 (1997).
26. S. A. Wainwright, *Nature* **224**, 777 (1969).
27. M. Kowalewski, P. H. Kelley, Eds., *The Fossil Record of Predation* (Paleontological Society Papers *8*) (Paleontological Society, Lancaster, PA, 2002).
28. R. K. Bambach, *Paleobiology* **19**, 372 (1993).
29. A. R. Palmer, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1379 (1992).
30. R. G. Maliva, R. Siever, *J. Geol.* **96**, 387 (1988).

31. I thank T. R. Waller, D. Jablonski, and J. Pojeta for stimulating discussions and D. Jablonski, M. Foote, M. Labarbera, T. A. Rothfus, M. Kowalewski, and two anonymous persons for helpful reviews. Analyses used the bivalve stratigraphic range data of J. J. Sepkoski Jr. as updated by D. Jablonski (version June 2004). Supported by NSF-EAR grant 0345897.

Supporting Online Material

www.sciencemag.org/cgi/content/full/307/5711/914/DC1
Materials and Methods
Figs. S1 to S4
Tables S1 to S5

22 October 2004; accepted 21 December 2004
10.1126/science.1106654

Two Abundant Bioaccumulated Halogenated Compounds Are Natural Products

Emma L. Teuten, Li Xu, Christopher M. Reddy*

Methoxylated polybrominated diphenyl ethers (MeO-PBDEs) have been found bioaccumulated in the tissues of a variety of aquatic animals and at concentrations comparable to those of anthropogenic halogenated organic compounds, including polychlorinated biphenyls (PCBs). The origin of the MeO-PBDEs has been uncertain; circumstantial evidence supports a natural and/or an industrial source. By analyzing the natural abundance radiocarbon content of two MeO-PBDEs isolated from a True's beaked whale (*Mesoplodon mirus*), we show that these compounds were naturally produced.

It has long been known that industrially produced halogenated organic compounds and their metabolites accumulate in human and other animal tissues. One example is polybrominated diphenyl ethers (PBDEs), used as flame retardants, which have emerged as a contaminant of concern because of the detection of increasing levels in human breast milk (*1*). Structurally related compounds, methoxylated polybrominated diphenyl ethers (MeO-PBDEs), have recently been identified in a variety of fish (*2*) and marine mammals (*3*); they form a new class of bioaccumulated compounds. The origin of these MeO-PBDEs is not immediately obvious. Many MeO-PBDEs are known natural products, isolated from sponges (*4*–*9*), algae (*10*), and acorn worms (*11*). However, their structural similarity to industrial PBDEs cannot be ignored. Industrially produced PBDEs lack a methoxy group, but this group could be introduced by

Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA.

*To whom correspondence should be addressed. E-mail: creddy@whoi.edu

direct methoxylation or a two-step hydroxylation followed by methylation, either of which is a viable biochemical transformation (*12*, *13*). Here, we show, by analysis of their radiocarbon (^{14}C) content, that two abundant MeO-PBDEs bioaccumulated in a North Atlantic True's beaked whale (*Mesoplodon mirus*) are natural products.

The MeO-PBDEs most frequently observed in animal tissue are 2-(2′,4′-dibromophenoxy)-3,5-dibromoanisole (MeO-BDE-47) and 2-(2′,4′-dibromophenoxy)-4,6-dibromoanisole (MeO-BDE-68) (*14*), whose structures are shown in Fig. 1A. The International Union of Pure and Applied Chemistry (IUPAC) system of labeling polychlorinated biphenyls (PCBs) has been used for naming the MeO-PBDEs. Circumstantial evidence supports bioaccumulation of MeO-PBDEs from both natural and industrial sources. MeO-BDE-68 has been observed at high concentrations in dolphins and other marine mammals in the same geographical location as sponges containing this compound (Queensland, Australia) (*15*). In contrast, several MeO-PBDEs were observed in fish from the Detroit River (*16*), a freshwater environment significantly affected by the highly industrialized area of the Great Lakes, and also in pike from several freshwater lakes in Sweden (*14*); no natural freshwater sources of these compounds are known. Additionally, the nonmethoxylated derivative of MeO-BDE-47 (BDE-47) is one of the most abundant industrial PBDEs in the environment (*2*).

Molecular-level ^{14}C analysis reliably determines whether halogenated organic compounds are derived from natural or industrial sources (*17*). With the exception of toxaphenes, industrial halogenated organic compounds are synthesized from petrochemicals, which are of geologic age and hence have no detectable ^{14}C ($\Delta^{14}C = -1000$ per mil; a notation representing a carbon source free of ^{14}C) (*17*). Conversely, compounds recently photosynthesized in the surface ocean are enriched ($\Delta^{14}C = +100 \pm 50$ per mil) depending on when and where the organism that produced it grew (*18*). Hence, observation of significant amounts of ^{14}C can be effectively used to distinguish a naturally produced compound from one manufactured industrially.

A True's beaked whale was chosen for this study because its blubber had a relatively high MeO-PBDE content, compared with some other marine mammals (*2*, *15*), and represented our first opportunity to obtain sufficient quantities of these compounds for ^{14}C analysis (a minimum of 100 μg). Analysis of the blubber (*19*) revealed that MeO-BDE-68 and MeO-BDE-47 are the fifth and sixth most abundant halogenated organic compounds (1.0 μg/g and 0.99 μg/g, respectively, that is per gram of lipid), present at concentrations comparable to typically observed halogenated organic compounds. The other halogenated organic compounds were the prominent pesticide metabolite 2,2-bis(*p*-chlorophenyl)-1,1-dichloroethene (DDE; 4.2 μg/g) and PCB-138, PCB-153, and PCB-180 (in the concentration range 1.0 to 1.2 μg/g). The MeO-PBDEs are more than 20 times as abundant as the most prevalent flame-retardant component, BDE-47 (0.038 μg/g). To ensure that enough of each MeO-PBDE was isolated to make a successful ^{14}C measurement, 10 kg of blubber was extracted. Upon isolation and purification, 0.71 mg of MeO-BDE-47 and 1.27 mg of MeO-BDE-68 were obtained in >99% purity (*19*). Half of each compound was set aside to confirm the structures, and the ^{14}C content was determined on the remaining portion (*19*). Modern $\Delta^{14}C$ values of +103 per mil and +119 per mil for MeO-BDE-47 and MeO-BDE-68, respectively, indicate that the bioaccumulated MeO-PBDEs were synthesized naturally and are not metabolites of flame-retardant PBDEs. These data, and other relevant $\Delta^{14}C$ results, are listed in Table 1 along with $\delta^{13}C$ data. The $\Delta^{14}C$ values obtained for the MeO-PBDEs are comparable with that determined for 2-(3′,5′-dibromo-2′-methoxyphenoxy)-3,5-dibromoanisole, a structurally similar compound isolated from the Pacific marine sponge *Phyllospongia foliascens* ($\Delta^{14}C = +73$ per mil). These observed $\Delta^{14}C$ values, in addition to the bulk organic carbon in the blubber (+92 per mil for the total lipid extract and +72 per mil for the whole blubber), are indicative of material that was recently photosynthesized in the surface ocean. Identification of an exact location, however, is complicated by geographical and temporal $\Delta^{14}C$ variations of the dissolved inorganic carbon. Contrasting these modern values, an essentially ^{14}C-free signal of –994 per mil for Bromkal 70-5DE, a commercial PBDE mixture, denotes a carbon skeleton derived from petrochemical feedstock.

It is difficult to discern the extent to which contaminants in the whale contributed to his death, although a chronic parasitic liver infection likely played a significant role (*20*).

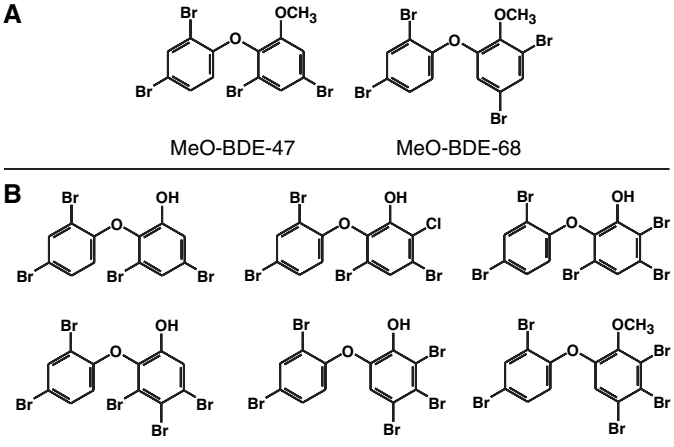


Fig. 1. Structures of PBDE derivatives. (**A**) MeO-PBDEs isolated from *M. mirus*. (**B**) Examples of PBDEs isolated from marine sponges that may be metabolized to MeO-BDE-47 or MeO-BDE-68, by methylation, dechlorination, and/or debromination.

Table 1. Radiocarbon ($\Delta^{14}C$) and $\delta^{13}C$ data for the isolated MeO-PBDEs and other relevant substances. MeO-BDE-47 and MeO-BDE-68 were isolated from the blubber of a True's beaked whale (*M. mirus*). 2-(3′,5′-Dibromo-2′-methoxy-phenoxy)-3,5-dibromoanisole was isolated from the marine sponge *P. foliascens*. Bromkal 70-5DE, Aroclor 1260, and DDT were industrially synthesized from petrochemical feedstock. Instrumental error in $\Delta^{14}C$ measurements is typically less than or equal to ±15 per mil for small samples (*30*). For both MeO-BDE-47 and MeO-BDE-68 the instrumental error was ±9 per mil. A single measurement was made on each sample; the precision of compound-specific $\Delta^{14}C$ data has been estimated as ~±20 per mil (*31*). Precision for analysis of a similar isolated halogenated organic compound was ±13 per mil (*18*). Johnson-Mathey 99.9999% graphite powder was employed as a process blank.

Substance	Source	$\Delta^{14}C$ (per mil)	$\delta^{13}C$ (per mil)	NOSAMS accession no.
MeO-BDE-47	*M. mirus*	+103	–22.2	OS-44624
MeO-BDE-68	*M. mirus*	+119	–24.2	OS-44623
Blubber total lipid extract	*M. mirus*	+92	–21.7	OS-44570
Whole blubber	*M. mirus*	+72	–21.7	OS-47041
2-(3′,5′-Dibromo-2′-methoxy-phenoxy)-3,5-dibromoanisole (*17*)	*P. foliascens*	+73	–35.9*	OS-30008
Bromkal 70-5DE (PBDE mixture) (*17*)	Ultra Scientific	–994	–26.7	OS-27545
Aroclor 1260 (PCB mixture) (*17*)	Ultra Scientific	–999	–26.7	OS-16065
DDT (*17*)	Ultra Scientific	–998	–28.3	OS-17853
Dissolved inorganic carbon in Atlantic Ocean (surface)† (*30*)		+100 ± 50		

*Isotopic fractionation may have occurred during isolation. †For the period 1970–1994.

Using U.S. Environmental Protection Agency EPI (Estimation Programs Interface) Suite software, we estimated octanol-water partitioning coefficients (K_{ow}) for the two MeO-PBDEs, log $K_{ow} \approx 6.85$; this result suggests a high propensity for bioaccumulation. The whale was found fatally stranded in False Cape State Park, Virginia, indicating that he was of the North Atlantic population of True's beaked whale, whose reported range extends from San Salvador Island, Bahamas, northward to St. Ann's Bay, Nova Scotia (*21*). To date, natural production of these compounds has not been reported in the Atlantic Ocean at any latitude, although this does not preclude their presence. However, both of these MeO-PBDEs are known natural products. MeO-BDE-47 has been isolated from *Dysidea herbacea* (*4*) and MeO-BDE-68 has been found in *D. dendyi* (*5*), both marine sponges from the Pacific and Indian Oceans. MeO-BDE-68 was also isolated from the Pacific green algae *Cladophora fascicularis* (*10*). It is possible that these MeO-PBDEs are also unidentified metabolites of other organisms; for example, *Cladophora* spp. is found throughout the world's oceans, and several other *Dysidea* spp. inhabit the Atlantic Ocean. PBDE derivatives isolated from *D. herbacea* are known to be synthesized by the symbiotic cyanobacteria *Oscillatoria spongeliae* (*22*); *Oscillatoria* spp. is distributed worldwide and presents another potential MeO-PBDE source.

The most likely form of exposure of the whale to these compounds is dietary, through consumption of squid, the major food for this species (*21*). Concentrations of MeO-PBDEs in squid have not been reported, so the transfer route is difficult to trace. The modern $\Delta^{14}C$ values for the MeO-PBDEs indicate a chemical half-life on the order of decades, as opposed to centuries or longer, as the observed $\Delta^{14}C$ values have not been significantly affected by the radioactive decay of ^{14}C. The results presented here show that the MeO-PBDEs were very recently naturally produced, in contrast with a previous attempt to use ^{14}C to elucidate the origin of 1,1′-dimethyl-3,3′,4,4′-tetrabromo-5,5′-dichloro-2,2′-bipyrrole, a bioaccumulated halogenated organic compound isolated from several marine mammals and birds (*18*). In this case, $\Delta^{14}C$ (–458 ± 13 per mil) revealed at least in part a natural source, but was considerably lower than values for recently synthesized natural products (*18*). More than 30 other naturally produced PBDE derivatives are known, yet they were not observed above trace concentrations in the True's beaked whale blubber. Compounds have been isolated from marine sponges with a hydroxyl group substituent (*6*, *7*), an additional methoxy group (*7*, *8*), or a greater level of bromination (*6*–*8*), in addition to mixed halogenated analogs also containing chlorine (*9*). The presence of parts per billion to parts per million concentrations of MeO-BDE-47 and MeO-BDE-68, but not other PBDE derivatives, in fish and marine mammals throughout the world might be explained if the former are relatively stable metabolites arising from many other structurally similar compounds. Hydroxylated polychlorinated diphenyl ethers are methylated by microorganisms (*13*), which suggests that the same transformation is plausible for polybrominated analogs. This may be a detoxification mechanism, as methyl derivatives exhibit less bioactivity than the corresponding alcohols (*8*, *23*). Because chlorinated diphenyl ethers are microbially metabolized more readily than brominated derivatives (*24*), it is feasible that chlorine is cleaved from mixed halogenated diphenyl ether natural products giving solely brominated compounds. Debromination of highly brominated PBDEs has also been shown, both under photolytic (*25*) and biological (*26*, *27*) conditions, and these reactions likely also apply to methoxylated derivatives. Reductive debromination of pentabrominated BDE-99 in the intestinal tract of carp generated BDE-47 (*26*), which has the same bromine substitution pattern as the bioaccumulated MeO-BDE-47. The structures of several compounds isolated from marine sponges that could viably be transformed to MeO-BDE-47 or MeO-BDE-68 under environmental conditions are shown in Fig. 1B. The stability of halogenated diphenyl ether natural products warrants further investigation and may be useful in understanding the fate and removal processes of industrial halogenated organic compounds in the environment.

That the bioaccumulated MeO-PBDEs are naturally produced does not mean they are benign. Bioactivity has been investigated only for MeO-BDE-68, which is reported to have antibacterial activity against *Escherichia coli*, *Bacillus subtilus*, and *Staphylococcus aureus*, in addition to potent anti-inflammatory activity (*10*). Because of the structural similarity of MeO-PBDEs with compounds known to induce adverse health effects, such as polychlorinated dibenzodioxins, PCBs, and PBDEs, other bioactivity is likely.

Certainly, these naturally produced halogenated organic compounds were in the environment and bioaccumulated many years before their industrial-scale manufacture and subsequent release, beginning in the 1920s (*28*). Evidence is accumulating to support the biotransformation of anthropogenic PBDEs (*26*, *27*). It is doubtful that these processes could have developed so rapidly in response to the anthropogenic input of halogenated organic compounds. Indeed, cytochrome P450 and other enzymes involved in the biotransformation and biodegradation of many persistent anthropogenic halogenated organic compounds are believed to have existed for millions of years, and they probably arose originally as a response to naturally produced compounds in the environment (*29*). Widespread dispersion of halogenated natural products with a tendency to bioaccumulate may have been important for such development.

Our findings reveal that naturally produced MeO-PBDEs can bioaccumulate in the food web. Based on these results, we see potential for using natural abundance ^{14}C measurements to uncover the relative contribution of industrial and natural sources of halogenated organic compounds to a variety of environmental media.

References and Notes

1. D. Meironyté, K. Norén, *J. Toxicol. Environ. Health* **58**, 329 (2003).
2. P. S. Haglund, D. R. Zook, H. Buser, J. Hu, *Environ. Sci. Technol.* **31**, 3281 (1997).
3. W. Vetter, E. Scholz, C. Gaus, J. F. Müller, D. Haynes, *Arch. Environ. Contam. Toxicol.* **41**, 221 (2001).
4. V. Anjaneyulu, K. Nageswara Rao, P. Radhika, M. Muralikrishna, J. D. Connolly, *Indian J. Chem.* **35B**, 89 (1996).
5. N. K. Utkina, V. A. Denisenko, M. V. Virovaya, O. V. Scholokova, N. G. Prokof'eva, *J. Nat. Prod.* **65**, 1213 (2002).
6. D. Handayani *et al.*, *J. Nat. Prod.* **60**, 1313 (1997).
7. B. Carté, D. J. Faulkner, *Tetrahedron* **37**, 2335 (1981).
8. H. Liu *et al.*, *J. Nat. Prod.* **67**, 472 (2004).
9. R. Capon, E. L. Ghisalberti, P. R. Jefferies, B. W. Skelton, A. H. White, *J. Chem. Soc. Perkin Trans.* **1**, 2464 (1981).
10. M. Kuniyoshi, K. Yamada, T. Higa, *Experientia* **41**, 523 (1985).
11. T. Higa, P. J. Scheuer, in *Marine Natural Products Chemistry*, D. J. Faulkner, W. H. Fenical, Eds. (Plenum Press, New York, 1977), pp. 35–43.
12. H. Hakk, R. J. Letcher, *Environ. Int.* **29**, 801 (2003).
13. K. Hundt *et al.*, *Appl. Environ. Microbiol.* **66**, 4157 (2000).
14. A. Kierkegaard *et al.*, *Environ. Pollut.* **130**, 187 (2004).
15. W. Vetter *et al.*, *Environ. Toxicol. Chem.* **21**, 2014 (2002).
16. R. J. Letcher *et al.*, *Organohal. Comp.* **61**, 29 (2003).
17. C. M. Reddy, L. Xu, T. I. Eglinton, J. P. Boon, D. J. Faulkner, *Environ. Pollut.* **120**, 163 (2002).
18. C. M. Reddy *et al.*, *Environ. Sci. Technol.* **38**, 1992 (2004).
19. Materials and Methods are available as supporting material on *Science* Online.
20. Virginia Marine Science Museum, "Necropsy examination report" (2003).
21. J. G. Mead, in *Handbook of Marine Mammals*, vol. 4, *River Dolphins and Larger Toothed Whales*, S. H. Ridgway, R. Harrison, Eds. (Academic Press, San Diego, 1989), pp. 349–30.
22. M. D. Unson, N. D. Holland, D. J. Faulkner, *Mar. Biol.* **119**, 1 (1994).
23. X. Fu *et al.*, *J. Nat. Prod.* **58**, 1384 (1995).
24. S. Schmidt *et al.*, *Appl. Environ. Microbiol.* **58**, 2744 (1992).
25. J. Eriksson, N. Green, G. Marsh, Å. Bergman, *Environ. Sci. Technol.* **38**, 3119 (2004).
26. H. M. Stapleton, R. J. Letcher, J. E. Baker, *Environ. Sci. Technol.* **38**, 1054 (2004).
27. A. Kierkegaard, L. Balk, U. Tjärnlund, C. A. de Wit, B. Jansson, *Environ. Sci. Technol.* **33**, 1612 (1999).
28. R. Stringer, P. Johnston, *Chlorine and the Environment: An Overview of the Chlorine Industry* (Kluwer Academic, Dordrecht, Netherlands, 2001).
29. J. J. Stegman, M. E. Hahn, in *Aquatic Toxicology: Molecular, Biochemical and Cellular Perspectives*, D. C. Malins, G. K. Ostrander, Eds. (Lewis Publishers, CRC Press, Boca Raton, FL, 1993), pp. 87–206.
30. A. Pearson, thesis, Woods Hole Oceanographic Institution and Massachusetts Institute of Technology (2000).

31. R. Nydal, *Radiocarbon* **42**, 81 (2000).
32. We thank the Virginia Marine Science Museum for providing us with *M. mirus* blubber; C. Johnson for conducting the experiments with nuclear magnetic resonance (NMR) and high-resolution mass spectrometry; R. Nelson for his assistance in the laboratory; M. Hahn, T. Eglinton, and J. Hayes for critiques of this manuscript; and Ö. Gustafsson for useful discussions. This work was supported by the National Science Foundation (CHE-0089172 and OCE-0221181), the Postdoctoral Scholar Program at Woods Hole Oceanographic Institution (WHOI) (with funding provided by the Henry and Camille Dreyfus Foundation Inc. and the J. Seward Johnson Fund, awarded to E.L.T.), and the WHOI Ocean Life Institute. This is WHOI contribution 11300.

Supporting Online Material
www.sciencemag.org/cgi/content/full/307/5711/917/DC1
Materials and Methods
References and Notes

28 October 2004; accepted 17 December 2004
11.1026/science.1106882

Deer Browsing and Population Viability of a Forest Understory Plant

James B. McGraw* and Mary Ann Furedi

American ginseng is the premier medicinal plant harvested from the wild in the United States. In this study, seven populations of ginseng plants were censused every 3 weeks during the growing season over 5 years to monitor deer browse and harvest and to project population growth and viability. The minimum viable population size was ~800 plants, a value greater than that of all populations currently being monitored. When simulated deer browsing rates were reduced 50% or more, population viability rose sharply. Without more effective deer population control, ginseng and many other valuable understory herbs are likely to become extinct in the coming century.

American ginseng (*Panax quinquefolius* L.) is a widespread but uncommon herbaceous understory plant of the U.S. eastern deciduous forest (*1*–*3*). The harvest of wild ginseng to supply the Asian market is economically and culturally important, particularly in central Appalachia. Though poorly quantified, many lines of evidence suggest that ginseng was more abundant in the presettlement forest than at present (*1*, *4*). Harvest figures from the 1800s suggest a three- to fourfold greater export of ginseng than currently occurs (*5*). Herbarium specimens show that the size of plants collected by botanists has shrunk significantly in the past century (*4*). Permanent land use conversion from forest due to farming, mining, and development has reduced the number of populations. Concern about the rarity of ginseng led to its listing on Appendix II of CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora) in 1973 (*6*).

The conservation of rare species benefits from demographic modeling based on accurate field data (*7*). Formal demographic analyses of ginseng were first performed for four populations at the northern margin of the distribution in southern Quebec (*8*, *9*). The resulting models showed that ginseng populations grow or decline slowly, with larger plants contributing most to population growth rate. The minimum viable population size (defined as having a 95% chance of persisting for 100 years) was estimated at 172 plants, a figure that raised serious concerns about the long-term future for ginseng, because most populations are smaller than this threshold. The extinction risk for ginseng near the range center in Appalachia may differ from that in Quebec, because populations are subject to the potential negative effects of harvest and deer browsing; however, there is also a greater frequency of near-optimal habitat.

In the present study, carried out over a period of 5 years (2000–2004), we censused seven natural populations of American ginseng in West Virginia in order to evaluate population viability (*10*). These populations varied in size and occurred over a range of elevations, aspects, and forest community types representative of the range center of ginseng (*1*). Transition matrices summarizing the fates of stage classes were formed from the census data for each pair of years (*10*, *11*).

Early in the censusing procedure, we noted that plants were being browsed by white-tailed deer (*Odocoileus virginianus* Zimm.); all foliage, and frequently all flowers and fruits, were removed (*10*). Annual mean browse rates varied from 19 to 42% across 4 years. This rate varied more widely across populations (10 to 63%) than among years. The browse rate was greater for adult plants (11 to 100%; weighted mean, 45%) than for seedlings and juveniles (2 to 64%; weighted mean, 21%). We also documented occasional legal and illegal harvest in the seven populations. Over four growing seasons, two of seven populations experienced harvest, and 0.45 to 3.04% of all monitored plants were harvested, a rate that is within the range of the overall rate observed in 36 monitored populations.

The dominant eigenvalue of the transition matrix yields the finite rate of increase (λ), a measure of annual growth rate for each population (*11*). λ varied from population to population and year to year. The geometric mean eigenvalue of the transition matrix for all populations (across years) was 0.973, representing an annual rate of decline of 2.7%. To examine the impact that deer were having on λ, we constructed equivalent matrices minus the deer-browsed plants. λ was estimated to be 1.021 for this "no browse" matrix, suggesting that populations would in fact grow slightly (2.1% per year) when the effect of browsing was removed. The increase in λ for the unbrowsed population was statistically significant (one-tailed paired t test; $P = 0.034$, $n = 4$ years). To further test the relationship of browse to λ, we regressed the deviation of λ from the annual mean λ versus browse rate for all population/year combinations. A significant negative relationship demonstrated that among-population variation in λ was explained in part by browse rate variation (slope = –0.318, $P = 0.022$, $r^2 = 0.19$).

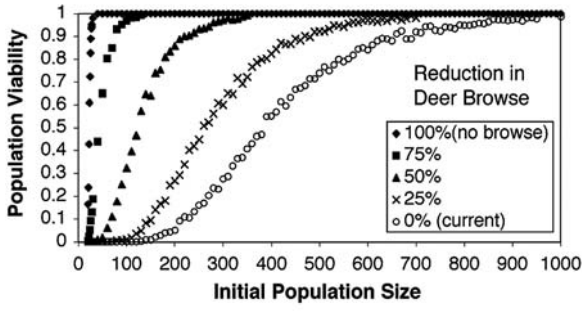


Fig. 1. Population viability as a function of initial population size at five levels of simulated browse.

Department of Biology, Post Office Box 6057, West Virginia University, Morgantown, WV 26506–6057, USA.

*To whom correspondence should be addressed. E-mail: jmcgraw@wvu.edu

31. R. Nydal, *Radiocarbon* **42**, 81 (2000).
32. We thank the Virginia Marine Science Museum for providing us with *M. mirus* blubber; C. Johnson for conducting the experiments with nuclear magnetic resonance (NMR) and high-resolution mass spectrometry; R. Nelson for his assistance in the laboratory; M. Hahn, T. Eglinton, and J. Hayes for critiques of this manuscript; and Ö. Gustafsson for useful discussions. This work was supported by the National Science Foundation (CHE-0089172 and OCE-0221181), the Postdoctoral Scholar Program at Woods Hole Oceanographic Institution (WHOI) (with funding provided by the Henry and Camille Dreyfus Foundation Inc. and the J. Seward Johnson Fund, awarded to E.L.T.), and the WHOI Ocean Life Institute. This is WHOI contribution 11300.

Supporting Online Material
www.sciencemag.org/cgi/content/full/307/5711/917/DC1
Materials and Methods
References and Notes

28 October 2004; accepted 17 December 2004
11.1026/science.1106882

Deer Browsing and Population Viability of a Forest Understory Plant

James B. McGraw* and Mary Ann Furedi

American ginseng is the premier medicinal plant harvested from the wild in the United States. In this study, seven populations of ginseng plants were censused every 3 weeks during the growing season over 5 years to monitor deer browse and harvest and to project population growth and viability. The minimum viable population size was ~800 plants, a value greater than that of all populations currently being monitored. When simulated deer browsing rates were reduced 50% or more, population viability rose sharply. Without more effective deer population control, ginseng and many other valuable understory herbs are likely to become extinct in the coming century.

American ginseng (*Panax quinquefolius* L.) is a widespread but uncommon herbaceous understory plant of the U.S. eastern deciduous forest (*1*–*3*). The harvest of wild ginseng to supply the Asian market is economically and culturally important, particularly in central Appalachia. Though poorly quantified, many lines of evidence suggest that ginseng was more abundant in the presettlement forest than at present (*1*, *4*). Harvest figures from the 1800s suggest a three- to fourfold greater export of ginseng than currently occurs (*5*). Herbarium specimens show that the size of plants collected by botanists has shrunk significantly in the past century (*4*). Permanent land use conversion from forest due to farming, mining, and development has reduced the number of populations. Concern about the rarity of ginseng led to its listing on Appendix II of CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora) in 1973 (*6*).

The conservation of rare species benefits from demographic modeling based on accurate field data (*7*). Formal demographic analyses of ginseng were first performed for four populations at the northern margin of the distribution in southern Quebec (*8*, *9*). The resulting models showed that ginseng populations grow or decline slowly, with larger plants contributing most to population growth rate. The minimum viable population size (defined as having a 95% chance of persisting for 100 years) was estimated at 172 plants, a figure that raised serious concerns about the long-term future for ginseng, because most populations are smaller than this threshold. The extinction risk for ginseng near the range center in Appalachia may differ from that in Quebec, because populations are subject to the potential negative effects of harvest and deer browsing; however, there is also a greater frequency of near-optimal habitat.

In the present study, carried out over a period of 5 years (2000–2004), we censused seven natural populations of American ginseng in West Virginia in order to evaluate population viability (*10*). These populations varied in size and occurred over a range of elevations, aspects, and forest community types representative of the range center of ginseng (*1*). Transition matrices summarizing the fates of stage classes were formed from the census data for each pair of years (*10*, *11*).

Early in the censusing procedure, we noted that plants were being browsed by white-tailed deer (*Odocoileus virginianus* Zimm.); all foliage, and frequently all flowers and fruits, were removed (*10*). Annual mean browse rates varied from 19 to 42% across 4 years. This rate varied more widely across populations (10 to 63%) than among years. The browse rate was greater for adult plants (11 to 100%; weighted mean, 45%) than for seedlings and juveniles (2 to 64%; weighted mean, 21%). We also documented occasional legal and illegal harvest in the seven populations. Over four growing seasons, two of seven populations experienced harvest, and 0.45 to 3.04% of all monitored plants were harvested, a rate that is within the range of the overall rate observed in 36 monitored populations.

The dominant eigenvalue of the transition matrix yields the finite rate of increase (λ), a measure of annual growth rate for each population (*11*). λ varied from population to population and year to year. The geometric mean eigenvalue of the transition matrix for all populations (across years) was 0.973, representing an annual rate of decline of 2.7%. To examine the impact that deer were having on λ, we constructed equivalent matrices minus the deer-browsed plants. λ was estimated to be 1.021 for this "no browse" matrix, suggesting that populations would in fact grow slightly (2.1% per year) when the effect of browsing was removed. The increase in λ for the unbrowsed population was statistically significant (one-tailed paired t test; $P = 0.034$, $n = 4$ years). To further test the relationship of browse to λ, we regressed the deviation of λ from the annual mean λ versus browse rate for all population/year combinations. A significant negative relationship demonstrated that among-population variation in λ was explained in part by browse rate variation (slope = –0.318, $P = 0.022$, $r^2 = 0.19$).

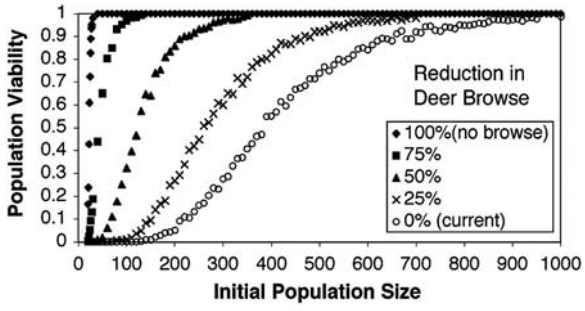


Fig. 1. Population viability as a function of initial population size at five levels of simulated browse.

Department of Biology, Post Office Box 6057, West Virginia University, Morgantown, WV 26506–6057, USA.

*To whom correspondence should be addressed. E-mail: jmcgraw@wvu.edu

Population viability analysis (PVA) provides a formal structure for making prognoses concerning extinction probabilities for populations (*10*–*12*). The procedure involves making stochastic population projections for a time frame of interest and determining whether the population goes extinct during that time (*10*). This process is repeated until an accurate estimate of viability (the chance of surviving that time frame) is reached. For our simulations, we chose a time frame of 100 years as the length of a simulation run, and we completed 1000 replicate runs to estimate the probability that a population would be viable. Population viability estimates vary with initial population size. Within a PVA, we can therefore also determine the minimum viable population (MVP) (*10*), in our case defined as the initial N corresponding to a 95% probability of persisting for 100 years. Although this is a liberal MVP criterion (in that it allows a 5% extinction risk), the existence of many ginseng populations in the wild means that more risk is likely to be acceptable in conserving any one population.

Population viability at current browse rates followed a sigmoidal curve (Fig. 1). Spline fitting of the curve suggested that second-order polynomial regression fits were adequate to describe the function locally. Therefore, to find the initial N corresponding to 95% population viability, we fit a second-order polynomial to the portion of the curve between viabilities of 0.92 and 0.98, solving the resulting quadratic equation for initial N at a viability of 0.95. We found an MVP of ~800 individuals (95% confidence limits; ~780 to 820) (*10*).

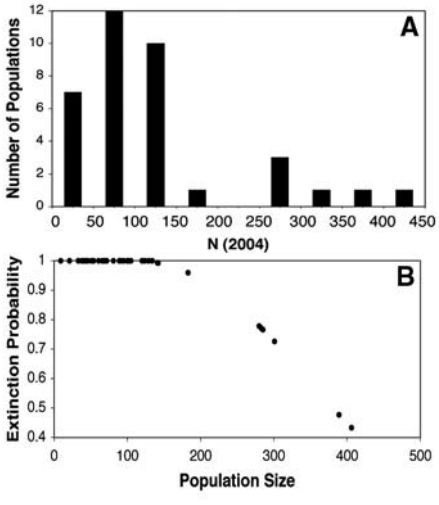


Fig. 2. (**A**) Frequency distribution of monitored population sizes in an eight-state region near the center of the range of ginseng and (**B**) estimated extinction probabilities for these same populations based on local polynomial fits to the viability versus N function in Fig. 1.

To put this new MVP estimate into perspective, Fig. 2A shows the distribution of population sizes we observed in an eight-state (Indiana, Kentucky, Maryland, New York, Ohio, Pennsylvania, Virginia, and West Virginia) demography study of 36 populations containing a total of 4448 plants. The median population size was 93 individuals, well below the MVP estimate. The maximum population size we observed was 406 individuals, which is only about half of the MVP, suggesting that by this relatively liberal definition, none of the 36 populations was viable. We know of only two natural populations in existence that exceed the MVP.

A plot of the chance of extinction within a century (1 minus viability) versus population size of these existing natural populations (Fig. 2B) shows that at current rates of browsing, all populations with $N < 143$ (29 of 36 of these natural populations) are projected to go extinct within a century (probability of extinction > 99%). The stochastic simulations showed that even the largest population had a 43.3% chance of going extinct before 100 years, a level considered unacceptable for conservation purposes. Acceptable risk is typically defined as 5% or less (*12*).

Because we observed high rates of deer browsing in the seven intensively studied populations described here, we investigated through simulation how reducing browse rates would affect population viability. This was done by randomly removing browsed individuals from the data set until the desired browse rate was achieved, then forming a new population projection matrix. After repeating this process 10 times for each year and finding the mean resulting matrix for each year, the entire PVA was repeated, and the resulting population viability functions are shown in Fig. 1 for simulated browse rates corresponding to browse rates reduced by 25, 50, 75, and 100% (no browse).

Deer reduced population viability and increased extinction risk. However, population viability increased rapidly with increasing initial N when browse rates were reduced in the simulations (Fig. 1). Concomitantly, MVP decreased as browse rates were reduced below current levels (Fig. 3). To place these results in the context of real populations, Fig. 4 shows the percent of populations we monitored in 2004 that would meet the MVP criteria as a function of deer browsing rate. A 50% reduction in deer browsing was required to achieve viability of any of the 36 populations we have censused, and viability monotonically increased as simulated herbivory rates were further reduced.

Any PVA should be cautiously interpreted and is only as good as the data used (*12*). Fortunately, ginseng's life cycle is amenable to accurate censusing, and its transition through biologically meaningful stages leads to a logical Lefkovitch matrix model formulation (*13*). Though never abundant, the existence of many natural ginseng populations justifies the liberal definition of population viability we adopted. By pooling data across populations, we may have underestimated the annual within-population variation in λ, which would in turn be likely to result in underestimating the risk of extinction. Therefore, in this respect, we believe that our analysis does not exaggerate the risks of extinction. Deforestation, the expansion of invasive species effects, further increases in deer populations, the introduction of disease from cultivated plants, and increases in the price of wild root could all worsen the prognosis. Alternatively, unpredictable positive effects might include the reintroduction of top carnivores, the spread of ungulate diseases, forest maturation, and the substitution in the marketplace of wild simulated ginseng (cultivated seeds grown untended under natural conditions) for plants from natural populations.

We conclude that current deer population densities in central Appalachia jeopardize the future of ginseng, as well as the culture of harvest and trade surrounding this important herb. Although demographic analyses of understory herbs are few, those that do exist suggest that deer represent a similar threat to many understory species (*14*, *15*), as well as tree seedlings and saplings (*16*–*18*). Indeed, deer have been identified as a keystone species that can have far-reaching direct and indirect effects on plant and animal communities (*16*, *18*, *19*).

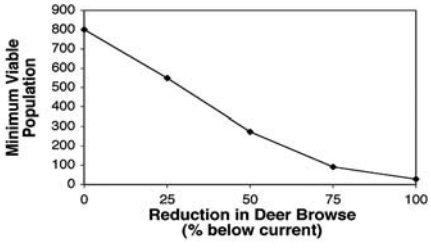


Fig. 3. Reduction in minimum viable population size as simulated deer browse rate is reduced.

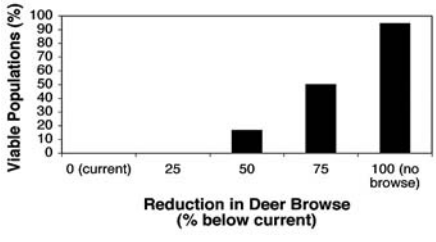


Fig. 4. Proportion of 36 monitored populations predicted to be viable as simulated deer browse rate is reduced from the current rate to no browse.

References and Notes

1. J. B. McGraw, S. M. Sanders, M. E. Van der Voort, *J. Torr. Bot. Soc.* **130**, 62 (2003).
2. R. C. Anderson, J. E. Armstrong, P. K. Benjamin, J. S. Fralish, *Am. Midl. Nat.* **129**, 357 (1993).
3. S. G. Carpenter, G. Cottam, *Can. J. Bot.* **60**, 2692 (1982).
4. J. B. McGraw, *Biol. Conserv.* **98**, 25 (2001).
5. A. W. Carlson, *Econ. Bot.* **40**, 233 (1986).
6. C. Robbins, *Conserv. Biol.* **14**, 1422 (2000).
7. D. W. Schemske *et al.*, *Ecology* **75**, 584 (1994).
8. D. Charron, D. Gagnon, *J. Ecol.* **79**, 431 (1991).
9. P. Nantel, D. Gagnon, A. Nault, *Conserv. Biol.* **10**, 608 (1996).
10. Information on materials and methods is available as supporting material on *Science* Online.
11. H. Caswell, *Matrix Population Models* (Sinauer, Sunderland, MA, ed. 2, 2001).
12. W. F. Morris, D. F. Doak, *Quantitative Conservation Biology* (Sinauer, Sunderland, MA, 2002).
13. L. P. Lefkovitch, *Biometrics* **21**, 1 (1965).
14. T. M. Knight, *Ecol. Appl.* **14**, 915 (2004).
15. D. J. Augustine, L. E. Frelich, *Conserv. Biol.* **12**, 995 (1998).
16. T. P. Rooney, *Forestry* **74**, 201 (2001).
17. F. L. Russell, D. B. Zippin, N. L. Fowler, *Am. Midl. Nat.* **146**, 1 (2001).
18. T. P. Rooney, D. M. Waller, *For. Ecol. Manag.* **181**, 165 (2003).
19. S. D. Côté, T. P. Rooney, J.-P. Tremblay, C. Dussault, D. M. Waller, *Annu. Rev. Ecol. Evol. Syst.* **35**, 113 (2004).

20. Supported by NSF grant DEB-0212411 to J.B.M. The authors thank W. Peterjohn, R. Landenberger, M. Van der Voort, G. Jochum, B. Bailey, and E. Mooney for comments and S. Lightner, M. Olive, R. May, G. Jochum, R. Kenyon, C. Packert, A. Lubbers, J. Wolf, and E. Mooney for assistance with fieldwork.

Supporting Online Material

www.sciencemag.org/cgi/content/full/307/5711/920/DC1
Materials and Methods
References

2 November 2004; accepted 14 December 2004
10.1126/science.1107036

Glycolipids as Receptors for *Bacillus thuringiensis* Crystal Toxin

Joel S. Griffitts,[1] Stuart M. Haslam,[2] Tinglu Yang,[3] Stephan F. Garczynski,[4] Barbara Mulloy,[5] Howard Morris,[6] Paul S. Cremer,[3] Anne Dell,[2] Michael J. Adang,[4] Raffi V. Aroian[1]*

The development of pest resistance threatens the effectiveness of *Bacillus thuringiensis* (Bt) toxins used in transgenic and organic farming. Here, we demonstrate that (i) the major mechanism for Bt toxin resistance in *Caenorhabditis elegans* entails a loss of glycolipid carbohydrates; (ii) Bt toxin directly and specifically binds glycolipids; and (iii) this binding is carbohydrate-dependent and relevant for toxin action in vivo. These carbohydrates contain the arthroseries core conserved in insects and nematodes but lacking in vertebrates. We present evidence that insect glycolipids are also receptors for Bt toxin.

The crystal (Cry) proteins produced by Bt are pore-forming toxins lethal to insects and nematodes but nontoxic to vertebrates (*1*, *2*). In 2002, more than 14 million hectares of transgenic corn and cotton crops that express Cry proteins were planted worldwide, making these crops safe from specific insect pests and simultaneously resulting in substantial decreases in hazardous chemical pesticide use (*3*, *4*). Cry proteins have now been shown to target nematodes as well, including the intestinal parasite *Nippostrongylus brasiliensis*, suggesting that Cry proteins may be used in the future to control parasitic nematodes of animals and plants (*5*). In the face of the enormous selective pressure generated by widespread use of Cry proteins in crops and organic farming, development of Cry toxin resistance among target populations is considered the major threat to their long-term use (*6*). The ability to detect resistance in the field, which is important for monitoring current resistance-management programs and making corrections before the resistance becomes a widespread problem, relies on molecular and genetic knowledge of the genes and pathways that give rise to resistance. Resistance can be mediated by multiple loci, the identities of which have remained largely elusive. To date, only insect cadherins, which serve as toxin receptors, have been definitively demonstrated to mutate to Cry toxin resistance (*7*, *8*). Other candidates for resistance alleles include a second Bt toxin–binding protein, aminopeptidase N, and a host protease required to process the Bt toxin (*9*, *10*). There are also a number of as yet unidentified loci that can mutate to Cry toxin resistance, including ones important for toxin binding (*11*, *12*).

Using forward genetics, we identified four genes (called *bre* genes for Bt toxin resistant) that mutate to Bt toxin resistance in the nematode *C. elegans* (*13*–*15*). Loss-of-function mutants in this pathway resist at least two Cry proteins, Cry5B, which targets nematodes (Fig. 1A), and Cry14A, which targets nematodes and insects (*13*, *14*). Cry5B and Cry14A are members of the main family of three-domain Bt toxins, which includes the commercially used Cry1, Cry2, and Cry3 toxins (*16*). The *bre* genes encode four glycosyltransferase proteins, act in a single pathway, and are required for the uptake of toxin into intestinal cells, suggesting that they might make a Bt toxin host cell receptor (*13*, *14*). Based on their in vitro activities, the BRE-3 and BRE-5 counterparts in *Drosophila*—EGGHEAD and BRAINIAC, respectively—have been suggested to synthesize the carbohydrate chains present on glycosphingolipids (*14*). We therefore hypothesized that the BRE enzymes might be involved in the biosynthesis of glycosphingolipids and that glycosphingolipids might be heretofore-unrecognized host cell receptors for Bt toxins.

To investigate these possibilities, lipids from wild-type and *bre* mutant animals were extracted, partitioned into two phases, resolved by thin-layer chromatography (TLC), and visualized with the orcinol reagent that stains carbohydrates (Fig. 1B). Wild-type animals contain multiple high-polarity glycolipid species (Fig. 1B, upper phase, components B to F). These glycolipids are ceramide-based (and hence glycosphingolipids) because the carbohydrates can be removed with leech ceramide glycanase (*17*). These upper phase glycolipids are completely absent in *bre-3*, *bre-4*, and *bre-5* mutant animals. In *bre-2* mutant animals, most (B, C, and F) but not all (D and E) upper phase components are missing. In contrast to what was seen in the upper phase, analysis of lower phase (presumably less complex) glycolipids from *bre-4* and *bre-5* mutant animals revealed the appearance of new glycolipid species (Fig. 1B), presumably each representing a different precursor that accumulates as a result of deficiencies in the biosynthetic pathway. Genetic epistasis allows us to infer that the BRE enzymes act in the following order in the synthesis of glycolipids: BRE-3, BRE-5, BRE-4, and lastly BRE-2 [supporting online material (SOM) text], in agreement with the known or proposed activities of these enzymes and the structures of their products. These data demonstrate that BRE enzymes are required to synthesize the carbohydrate chain of glycolipids. The lack of observable defects in protein-linked carbohydrates based on mass spectrometry analysis of N- and O-linked glycans from *bre-3* animals suggests that BRE-3 is not involved in the synthesis of glycoproteins (fig. S5 and table S5). These data and the fact that linkages dependent on *bre-3* and *bre*-5 have been found only in glycolipids indicate that glycolipids and not

[1]Section of Cell and Developmental Biology, University of California, San Diego, La Jolla, CA 92093–0349, USA. [2]Department of Biological Sciences, Imperial College London, London, SW7 2AZ, UK. [3]Department of Chemistry, Texas A&M University, College Station, TX 77843, USA. [4]Department of Entomology, University of Georgia, Athens, GA 30602–2603, USA. [5]Laboratory for Molecular Structure, National Institute for Biological Standards and Control, Blanche Lane, South Mimms, Potters Bar, Hertfordshire EN6 3QG, UK. [6]M-SCAN Mass Spectrometry Research and Training Centre, Silwood Park, Ascot, Berkshire SL5 7PZ, UK.

*To whom correspondence should be addressed. E-mail: raroian@ucsd.edu

References and Notes

1. J. B. McGraw, S. M. Sanders, M. E. Van der Voort, *J. Torr. Bot. Soc.* **130**, 62 (2003).
2. R. C. Anderson, J. E. Armstrong, P. K. Benjamin, J. S. Fralish, *Am. Midl. Nat.* **129**, 357 (1993).
3. S. G. Carpenter, G. Cottam, *Can. J. Bot.* **60**, 2692 (1982).
4. J. B. McGraw, *Biol. Conserv.* **98**, 25 (2001).
5. A. W. Carlson, *Econ. Bot.* **40**, 233 (1986).
6. C. Robbins, *Conserv. Biol.* **14**, 1422 (2000).
7. D. W. Schemske *et al.*, *Ecology* **75**, 584 (1994).
8. D. Charron, D. Gagnon, *J. Ecol.* **79**, 431 (1991).
9. P. Nantel, D. Gagnon, A. Nault, *Conserv. Biol.* **10**, 608 (1996).
10. Information on materials and methods is available as supporting material on *Science* Online.
11. H. Caswell, *Matrix Population Models* (Sinauer, Sunderland, MA, ed. 2, 2001).
12. W. F. Morris, D. F. Doak, *Quantitative Conservation Biology* (Sinauer, Sunderland, MA, 2002).
13. L. P. Lefkovitch, *Biometrics* **21**, 1 (1965).
14. T. M. Knight, *Ecol. Appl.* **14**, 915 (2004).
15. D. J. Augustine, L. E. Frelich, *Conserv. Biol.* **12**, 995 (1998).
16. T. P. Rooney, *Forestry* **74**, 201 (2001).
17. F. L. Russell, D. B. Zippin, N. L. Fowler, *Am. Midl. Nat.* **146**, 1 (2001).
18. T. P. Rooney, D. M. Waller, *For. Ecol. Manag.* **181**, 165 (2003).
19. S. D. Côté, T. P. Rooney, J.-P. Tremblay, C. Dussault, D. M. Waller, *Annu. Rev. Ecol. Evol. Syst.* **35**, 113 (2004).

20. Supported by NSF grant DEB-0212411 to J.B.M. The authors thank W. Peterjohn, R. Landenberger, M. Van der Voort, G. Jochum, B. Bailey, and E. Mooney for comments and S. Lightner, M. Olive, R. May, G. Jochum, R. Kenyon, C. Packert, A. Lubbers, J. Wolf, and E. Mooney for assistance with fieldwork.

Supporting Online Material

www.sciencemag.org/cgi/content/full/307/5711/920/DC1
Materials and Methods
References

2 November 2004; accepted 14 December 2004
10.1126/science.1107036

Glycolipids as Receptors for *Bacillus thuringiensis* Crystal Toxin

Joel S. Griffitts,[1] Stuart M. Haslam,[2] Tinglu Yang,[3] Stephan F. Garczynski,[4] Barbara Mulloy,[5] Howard Morris,[6] Paul S. Cremer,[3] Anne Dell,[2] Michael J. Adang,[4] Raffi V. Aroian[1]*

The development of pest resistance threatens the effectiveness of *Bacillus thuringiensis* (Bt) toxins used in transgenic and organic farming. Here, we demonstrate that (i) the major mechanism for Bt toxin resistance in *Caenorhabditis elegans* entails a loss of glycolipid carbohydrates; (ii) Bt toxin directly and specifically binds glycolipids; and (iii) this binding is carbohydrate-dependent and relevant for toxin action in vivo. These carbohydrates contain the arthroseries core conserved in insects and nematodes but lacking in vertebrates. We present evidence that insect glycolipids are also receptors for Bt toxin.

The crystal (Cry) proteins produced by Bt are pore-forming toxins lethal to insects and nematodes but nontoxic to vertebrates (*1*, *2*). In 2002, more than 14 million hectares of transgenic corn and cotton crops that express Cry proteins were planted worldwide, making these crops safe from specific insect pests and simultaneously resulting in substantial decreases in hazardous chemical pesticide use (*3*, *4*). Cry proteins have now been shown to target nematodes as well, including the intestinal parasite *Nippostrongylus brasiliensis*, suggesting that Cry proteins may be used in the future to control parasitic nematodes of animals and plants (*5*). In the face of the enormous selective pressure generated by widespread use of Cry proteins in crops and organic farming, development of Cry toxin resistance among target populations is considered the major threat to their long-term use (*6*). The ability to detect resistance in the field, which is important for monitoring current resistance-management programs and making corrections before the resistance becomes a widespread problem, relies on molecular and genetic knowledge of the genes and pathways that give rise to resistance. Resistance can be mediated by multiple loci, the identities of which have remained largely elusive. To date, only insect cadherins, which serve as toxin receptors, have been definitively demonstrated to mutate to Cry toxin resistance (*7*, *8*). Other candidates for resistance alleles include a second Bt toxin–binding protein, aminopeptidase N, and a host protease required to process the Bt toxin (*9*, *10*). There are also a number of as yet unidentified loci that can mutate to Cry toxin resistance, including ones important for toxin binding (*11*, *12*).

Using forward genetics, we identified four genes (called *bre* genes for Bt toxin resistant) that mutate to Bt toxin resistance in the nematode *C. elegans* (*13*–*15*). Loss-of-function mutants in this pathway resist at least two Cry proteins, Cry5B, which targets nematodes (Fig. 1A), and Cry14A, which targets nematodes and insects (*13*, *14*). Cry5B and Cry14A are members of the main family of three-domain Bt toxins, which includes the commercially used Cry1, Cry2, and Cry3 toxins (*16*). The *bre* genes encode four glycosyltransferase proteins, act in a single pathway, and are required for the uptake of toxin into intestinal cells, suggesting that they might make a Bt toxin host cell receptor (*13*, *14*). Based on their in vitro activities, the BRE-3 and BRE-5 counterparts in *Drosophila*—EGGHEAD and BRAINIAC, respectively—have been suggested to synthesize the carbohydrate chains present on glycosphingolipids (*14*). We therefore hypothesized that the BRE enzymes might be involved in the biosynthesis of glycosphingolipids and that glycosphingolipids might be heretofore-unrecognized host cell receptors for Bt toxins.

To investigate these possibilities, lipids from wild-type and *bre* mutant animals were extracted, partitioned into two phases, resolved by thin-layer chromatography (TLC), and visualized with the orcinol reagent that stains carbohydrates (Fig. 1B). Wild-type animals contain multiple high-polarity glycolipid species (Fig. 1B, upper phase, components B to F). These glycolipids are ceramide-based (and hence glycosphingolipids) because the carbohydrates can be removed with leech ceramide glycanase (*17*). These upper phase glycolipids are completely absent in *bre-3*, *bre-4*, and *bre-5* mutant animals. In *bre-2* mutant animals, most (B, C, and F) but not all (D and E) upper phase components are missing. In contrast to what was seen in the upper phase, analysis of lower phase (presumably less complex) glycolipids from *bre-4* and *bre-5* mutant animals revealed the appearance of new glycolipid species (Fig. 1B), presumably each representing a different precursor that accumulates as a result of deficiencies in the biosynthetic pathway. Genetic epistasis allows us to infer that the BRE enzymes act in the following order in the synthesis of glycolipids: BRE-3, BRE-5, BRE-4, and lastly BRE-2 [supporting online material (SOM) text], in agreement with the known or proposed activities of these enzymes and the structures of their products. These data demonstrate that BRE enzymes are required to synthesize the carbohydrate chain of glycolipids. The lack of observable defects in protein-linked carbohydrates based on mass spectrometry analysis of N- and O-linked glycans from *bre-3* animals suggests that BRE-3 is not involved in the synthesis of glycoproteins (fig. S5 and table S5). These data and the fact that linkages dependent on *bre-3* and *bre*-5 have been found only in glycolipids indicate that glycolipids and not

[1]Section of Cell and Developmental Biology, University of California, San Diego, La Jolla, CA 92093–0349, USA. [2]Department of Biological Sciences, Imperial College London, London, SW7 2AZ, UK. [3]Department of Chemistry, Texas A&M University, College Station, TX 77843, USA. [4]Department of Entomology, University of Georgia, Athens, GA 30602–2603, USA. [5]Laboratory for Molecular Structure, National Institute for Biological Standards and Control, Blanche Lane, South Mimms, Potters Bar, Hertfordshire EN6 3QG, UK. [6]M-SCAN Mass Spectrometry Research and Training Centre, Silwood Park, Ascot, Berkshire SL5 7PZ, UK.

*To whom correspondence should be addressed. E-mail: raroian@ucsd.edu

glycoproteins are important for *bre*-mediated Bt toxin susceptibility.

We next tested whether Cry5B can directly bind glycolipids. An overlay technique was used in which crude *C. elegans* glycolipids were fixed in place on TLC plates and then incubated in an aqueous solution of activated, biotinylated Cry5B. After washing away unbound toxin, Cry5B bound to glycolipids was detected by enzyme-linked biotin detection. We found that Cry5B is able to bind to a number of glycosphingolipid species, namely components B, C, E, F, and other minor species (Fig. 2A). Specificity of binding is demonstrated by our observations that neither glycolipid species D (Fig. 2A, lanes 2 and 3) nor the simple glycolipids that accumulate in *bre-4* and *bre-5* mutants (*17*) nor mammalian glycolipid standards (Fig. 2A, lanes 7 and 8) bind Cry5B. As predicted for our resistant mutants, Cry5B-binding glycolipids are missing in *bre-3*, *bre-4*, and *bre-5* mutant animals (Fig. 2A, lanes 4 to 6), and all but one is missing in *bre-2* mutant animals (band E; Fig. 2A, lane 3). Because *bre-2* mutant animals are as resistant as the other mutants (*14*, *15*), expression of band E must not be sufficient for intoxication, perhaps because that glycolipid species is not expressed on the apical surface of intestinal cells.

The binding of Cry5B to purified *C. elegans* glycolipids was confirmed in supported lipid bilayers with microfluidic methods (*18*). Glycolipid component B was purified and incorporated into phosphocholine liposomes at 0.35 mole percent. These liposomes were allowed to form a continuous bilayer in hydrophilic microchannels, and the binding of fluorescently labeled Cry5B was evaluated with total internal reflection fluorescence microscopy. Cry5B binding to component B occurs in a saturable, dose-dependent manner and exhibits an apparent dissociation constant, K_d (±SD), of 0.73 ± 0.06 μM at the particular ligand density tested (Fig. 2B). This K_d falls near the low end of the range observed for many protein lectin–carbohydrate interactions (*19*). No specific binding in the absence of component B was detected. Thus, *C. elegans* glycolipid component B is sufficient to generate specific binding sites for Cry5B toxin in lipid bilayers.

We determined the chemical structures of components B, C, D, and E (Fig. 2C, figs. S1 to S4, and tables S1 to S4). All of these structures contain the core tetrasaccharide *N*-acetylgalactosamine (GalNAc) β1–4 *N*-acetylglucosamine (GlcNAc) β1–3 mannose (Man) β1–4 glucose (Glc), which is an invertebrate-specific glycolipid signature conserved between nematodes and insects but lacking in vertebrates (*20*). Components D and E correspond to the previously described glycosphingolipid structures Nz2 and Nz3, respectively (*21*); the structures of components B and C were previously uncharacterized. The role of BRE-3, BRE-4, and BRE-5 in the synthesis of these structures can be assigned (Fig. 2C) on the basis of epistasis and the predicted or demonstrated biochemical activities of these enzymes (*14*, *22*). We propose that BRE-2 initiates the synthesis of the branched moiety that distinguishes components B and C from D and E (SOM text).

To evaluate the carbohydrate dependence of Cry5B binding to glycolipids, we examined the ability of simple sugars to inhibit the

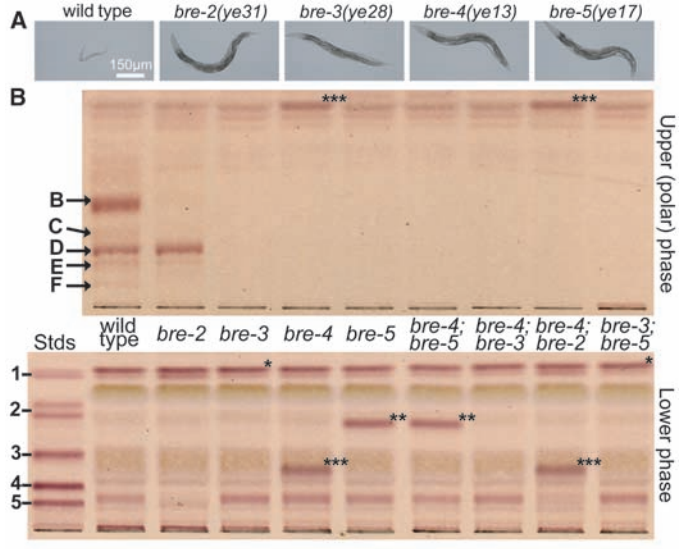


Fig. 1. Bt toxin–resistant animals are defective in glycolipid synthesis. (**A**) Resistance of *bre* mutants to Cry5B is shown after a 3-day exposure of L1 larvae to Cry5B. The alleles shown are used throughout this study. (**B**) Glycolipids were resolved by TLC and stained with orcinol and sulfuric acid. The origin is always at the bottom. Glycolipids stain reddish-brown and contaminating lipids stain yellow or gray. Lipid samples are derived from wild-type animals or from *bre* single- and double-mutant animals, as indicated, and separated as described in (*28*). Glycolipid components are designated as B to F. Numbers to the left denote the saccharide length of glycolipid standards (Stds) at those positions. Asterisks mark intermediates in the pathway where *bre* mutants are presumed to be blocked; the number of asterisks indicates the number of saccharides estimated to exist on the headgroup.

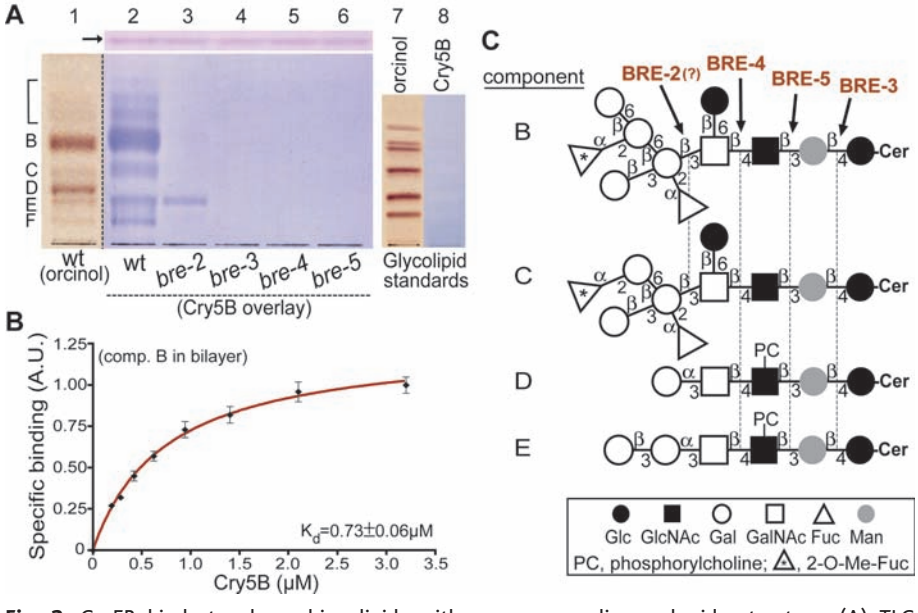


Fig. 2. Cry5B binds to glycosphingolipids with a common oligosaccharide structure. (**A**) TLC overlay reveals the Cry5B-glycolipid interaction. Lane 1, upper phase lipids from wild-type (wt) animals stained with orcinol; lanes 2 to 6, lipids from wild-type and mutant animals overlayed with biotinylated Cry5B; lanes 7 and 8, neutral glycolipid standards stained with orcinol (lane 7) but failed to bind biotinylated Cry5B (lane 8). Bracket indicates additional *bre*-dependent, toxin-binding glycolipids; arrow indicates toxin-independent coloration of a lipid contaminant to verify equal loading. (**B**) Specific binding of Cry5B to glycolipid component B (comp. B) incorporated into a bilayer. Cry5B labeled with Alexa 594 was injected over the membrane at various concentrations and the fluorescence was measured and normalized to data from control bilayers. Data points represent mean specific binding from four experiments; error bars denote standard deviation from the mean. A.U., arbitrary units. (**C**) *bre*-dependent glycolipids are complex ceramide (Cer)–linked oligosaccharides based on a common oligosaccharide core. Structural analysis was performed (*28*), and the glycosidic linkages proposed to be catalyzed by the BRE enzymes are indicated by arrows and dashed lines. Fuc, fucose.

binding of toxin to glycolipids. Glycolipid component B was purified and immobilized in polystyrene wells and then probed with biotinylated Cry5B in the absence and presence of various monosaccharides (Fig. 3A). Galactose is the most potent of the monosaccharide inhibitors, exerting 92 ± 2% binding inhibition at 15 mM. GalNAc also had a significant effect. The galactose analog β-methylgalactoside (conferring 92 ± 4% inhibition at 3 mM) was more inhibitory than the related compound α-methylgalactoside; β-methylglucoside was noninhibitory (Fig. 3B). β-galactose–mediated inhibition also occurred in our microfluidic lipid bilayer system (*17*). Galactose inhibits Cry5B binding to the entire *bre*-dependent glycolipid series in overlay assays (Fig. 3C), suggesting a common galactose-dependent binding mechanism. These data confirm that carbohydrates are key mediators of Cry5B binding to glycolipids and point to the β-galactose–rich terminus of these receptors as an important binding epitope.

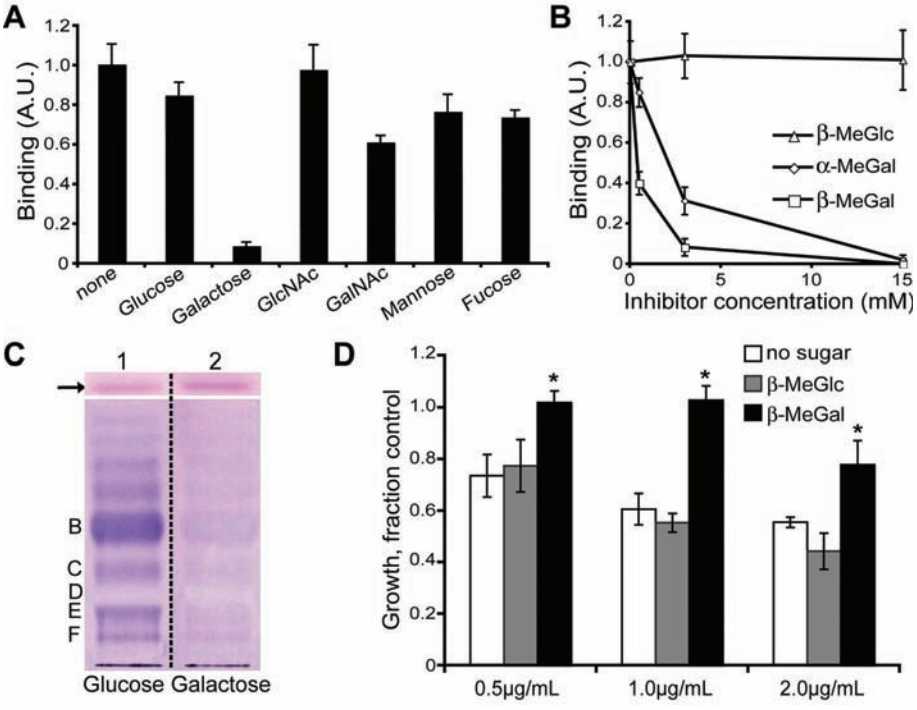


Fig. 3. β-galactose functions in Cry5B-glycolipid binding in vitro and intoxication in vivo. (**A**) Component B was immobilized in polystyrene wells and probed with biotinylated Cry5B in the absence or presence of 15 mM of the sugars indicated. Binding units reflect OD405 units with background subtracted. A.U., arbitrary units. (**B**) Comparison of binding inhibition imposed by anomeric monosaccharide derivatives. Binding was evaluated with the same assay as in (A) but with varying concentrations of inhibitors. β-MeGlc, β-methylglucoside; α-MeGal, α-methylgalactoside; β-MeGal, β-methylgalactoside. (**C**) Galactose inhibits Cry5B binding to all *bre*-dependent glycolipids. Lane 1, binding of biotinylated Cry5B in the presence of 100 mM glucose; lane 2, binding in the presence of 100 mM galactose. Arrow indicates toxin-independent coloration of a lipid contaminant to verify equal loading. (**D**) β-methylgalactoside protects *C. elegans* from Cry5B intoxication. Nematode growth in three doses of Cry5B protoxin is expressed after normalization to no-toxin controls. The test compounds were applied at 15 mM. All values are derived from three independent experiments each testing at least 20 animals. *, $P < 0.05$ versus control β-methylglucoside treatment. Error bars in (A), (B), and (D) denote standard deviation from the mean.

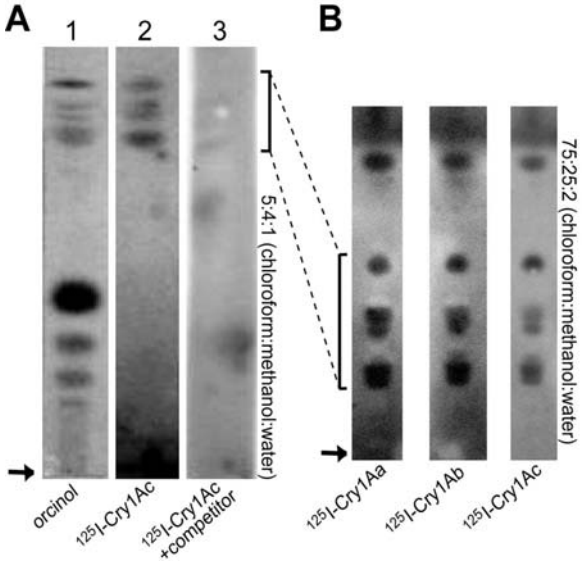


Fig. 4. The insecticidal toxins Cry1Aa, Cry1Ab, and Cry1Ac bind to *M. sexta* glycolipids. (**A**) In each lane, 20 μg of midgut-derived lipids was resolved in 5:4:1 (chloroform:methanol:water). Glycolipids were detected with orcinol (lane 1) or ^{125}I-Cry1Ac overlay (lanes 2 and 3). The experiment in lane 3 included 100 nM of cold Cry1Ac competitor. (**B**) Cry1Aa and Cry1Ab bind to the same *M. sexta* glycolipids as Cry1Ac. Midgut-derived lipids were resolved in 75:25:2 (chloroform:methanol:water), conditions that better separate the cluster of Cry1A-binding glycolipids, and probed with ^{125}I-Cry1Aa, ^{125}I-Cry1Ab, and ^{125}I-Cry1Ac. Brackets indicate Cry1A-binding glycolipids. Arrows indicate origins.

An in vivo prediction from these results is that β-methylgalactoside fed to *C. elegans* should provide an antidote to Cry5B toxin by competing with intestinal glycolipids for toxin binding. *C. elegans* hermaphrodites were fed doses of Cry5B that moderately inhibit nematode growth along with β-methylgalactoside, β-methylglucoside, or no exogenous carbohydrate. Neither of the carbohydrate treatments resulted in major growth differences in the absence of Cry5B (*17*). In the presence of Cry5B, β-methylgalactoside specifically protected animals at the toxin doses tested (Fig. 3D). Control glucoside-treated animals exhibited no protection from the toxin. Thus, the same treatment that directly interferes with the Cry5B-glycolipid interaction also specifically diminishes Cry5B toxicity, confirming the functional importance of these carbohydrate receptors to Cry toxin function in vivo.

Considering the substantial conservation of *bre*-dependent glycolipids in nematodes and insects and the conservation of Cry toxin structures (including lectin-like domains), it seems likely that insecticidal Cry toxin activity is also modulated by glycolipid host cell receptors. Consistent with this, we found that Cry1Ac toxin binds to glycolipids extracted from the midguts of the tobacco hornworm, *Manduca sexta* (Fig. 4A) (*23*). Competition of binding with unlabeled Cry1Ac indicates that binding is specific (Fig. 4A). Furthermore, Cry1Aa and Cry1Ab toxins bind the same *M. sexta* glycolipids as do Cry1Ac, consistent with glycolipids serving as general host cell receptors for these toxins (Fig. 4B).

Previously, glycolipid levels have been found to be substantially reduced in a Cry1Ac-resistant *Plutella xylostella* strain (*24*). It was proposed in the same report that alterations in glycolipids are involved in the evolution of *P. xylostella* resistance to Cry1Ac, although possible glycolipid receptor functions were not discussed. In a second study, a capacity for nonpurified *Bt* kurstaki toxins to bind to insect glycolipids was shown, but it was postulated that the in vivo toxin receptors were other glycoconjugates, such as glycoproteins (*25*). In addition, we have shown that Cry14A, a toxin

active against both nematodes and insects, requires the *bre* pathway for full activity against *C. elegans* (*13*, *14*). Taken together, our data and these studies suggest that both nematicidal and insecticidal three-domain Bt toxins use invertebrate glycolipids as host cell receptors and that the loss of glycolipid receptors represents an important mechanism for Bt toxin resistance. The ease with which glycolipids can be isolated and analyzed suggests that it will be feasible to monitor glycolipid-mediated resistance in field and laboratory populations of insects and nematodes.

Glycolipids, however, are not the only class of Bt toxin host cell receptors. For insects, high-affinity protein receptors, such as insect cadherins and aminopeptidases, have been shown to play functional roles as Cry1 receptors, although correlating protein receptor defects with binding defects has not always been simple. For example, in at least two cases, Cry1Ac has been found to bind specifically to membranes of cadherin receptor mutants that are Cry1Ac resistant (*26*, *27*), leading to the hypothesis that a multistep binding process involving multiple receptors is required for proper pore formation. For nematodes, our data suggest that although *bre*-dependent glycolipids are important for Cry14A function, there is likely another binding factor involved in Cry14A toxicity (*14*). We hypothesize that glycolipid and protein receptors may both play a role, sequentially or simultaneously, in positioning Bt toxins appropriately at the bilayer or in inserting toxins into the bilayer.

Mammalian cells do not bind three-domain Bt Cry toxins (*2*), and the results presented here provide a plausible molecular basis for the lack of toxicity of Cry toxins toward vertebrates. Vertebrates lack arthroseries glycolipids, which contain the conserved invertebrate-specific core tetrasaccharide GalNAcβ1-4GlcNAcβ1-3Manβ1-4Glc that is synthesized by the BRE pathway. Although the β-linked galactose important for Cry5B binding is not present in this core sequence, our unpublished data indicate that the intact receptor is, by three orders of magnitude, a better competitive inhibitor than β-methylgalactoside. Thus, higher-order structure is likely important for binding. We hypothesize that three-domain Bt Cry toxins evolved to at least partly recognize the core arthroseries tetrasaccharide and thus target the invertebrates, nematodes, and insects that synthesize these molecules.

The high degree of conservation between glycolipids present in *C. elegans* and in the human parasitic nematodes *Ascaris suum* and *Onchocerca volvolus*, which are phylogenetically divergent from *C. elegans*, suggests that most, if not all, nematodes will be susceptible to Cry5B toxin. All the nematodes we have tested to date are susceptible to Cry5B, and the one animal parasite we have tested has been shown to succumb to both Cry5B and Cry14A (*5*). Given these data and evidence that Cry5B and Cry14A toxin use an invertebrate-type glycolipid as their receptor, these Cry proteins hold great promise for one day safely targeting nematode pests of animals and plants.

References and Notes

1. R. A. de Maagd, A. Bravo, N. Crickmore, *Trends Genet.* **17**, 193 (2001).
2. F. S. Betz, B. G. Hammond, R. L. Fuchs, *Regul. Toxicol. Pharmacol.* **32**, 156 (2000).
3. M. Qaim, D. Zilberman, *Science* **299**, 900 (2003).
4. C. James, *ISAAA Briefs* **27**, 1 (2002).
5. J. Z. Wei *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 2760 (2003).
6. F. Gould, *Annu. Rev. Entomol.* **43**, 701 (1998).
7. L. J. Gahan, F. Gould, D. G. Heckel, *Science* **293**, 857 (2001).
8. S. Morin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5004 (2003).
9. R. Rajagopal, S. Sivakumar, N. Agrawal, P. Malhotra, R. K. Bhatnagar, *J. Biol. Chem.* **277**, 46849 (2002).
10. B. Oppert, K. J. Kramer, R. W. Beeman, D. Johnson, W. H. McGaughey, *J. Biol. Chem.* **272**, 23473 (1997).
11. D. G. Heckel, L. C. Gahan, F. Gould, A. Anderson, *J. Econ. Entomol.* **90**, 75 (1997).
12. J. L. Jurat-Fuentes, F. L. Gould, M. J. Adang, *Appl. Environ. Microbiol.* **68**, 5711 (2002).
13. J. S. Griffitts, J. L. Whitacre, D. E. Stevens, R. V. Aroian, *Science* **293**, 860 (2001).
14. J. S. Griffitts *et al.*, *J. Biol. Chem.* **278**, 45594 (2003).
15. L. D. Marroquin, D. Elyassnia, J. S. Griffitts, J. S. Feitelson, R. V. Aroian, *Genetics* **155**, 1693 (2000).
16. R. A. de Maagd, A. Bravo, C. Berry, N. Crickmore, H. E. Schnepf, *Annu. Rev. Genet.* **37**, 409 (2003).
17. J. S. Griffitts *et al.*, data not shown.
18. T. Yang, O. K. Baryshnikova, H. Mao, M. A. Holden, P. S. Cremer, *J. Am. Chem. Soc.* **125**, 4779 (2003).
19. B. E. Collins, J. C. Paulson, *Curr. Opin. Chem. Biol.* **8**, 617 (2004).
20. J. Mucha *et al.*, *Biochem. J.* **382**, 67 (2004).
21. S. Gerdt, R. D. Dennis, G. Borgonie, R. Schnabel, R. Geyer, *Eur. J. Biochem.* **266**, 952 (1999).
22. Z. S. Kawar, I. Van Die, R. D. Cummings, *J. Biol. Chem.* **277**, 34924 (2002).
23. S. F. Garczynki, M. J. Adang, in *Entomopathogenic Bacteria: From Laboratory to Field Application*, J.-F. Charles, A. Delecluse, C. Nielsen-LeRoux, Eds. (Kluwer, Netherlands, 2000), pp. 181–197.
24. N. S. Kumaraswami *et al.*, *Comp. Biochem. Physiol. B* **129**, 173 (2001).
25. R. D. Dennis, H. Wiegandt, D. Haustein, B. H. Knowles, D. J. Ellar, *Biomed. Chromatogr.* **1**, 31 (1986).
26. M. K. Lee, F. Rajamohan, F. Gould, D. H. Dean, *Appl. Environ. Microbiol.* **61**, 3836 (1995).
27. J. Gonzalez-Cabrera, B. Escriche, B. E. Tabashnik, J. Ferre, *Insect Biochem. Mol. Biol.* **33**, 929 (2003).
28. Materials and methods are available as supporting material on *Science* Online.

29. We thank R. Schnaar for technical advice on glycolipids, B. Hayes and N. Preece of the UCSD Glycotechnology Core for help with glycan analysis, T. Huxford for assistance with toxin purification, D. Huffman for comments, and members of the R.V.A. laboratory for stimulating discussions. This work was supported by NSF grant MCB-9983013 and grants from the Burroughs–Wellcome Foundation and the Beckman Foundation (to R.V.A.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/307/5711/922/DC1
Materials and Methods
SOM Text
Figs. S1 to S5
Tables S1 to S5
References

24 August 2004; accepted 19 November 2004
10.1126/science.1104444

Lymphotoxin-Mediated Regulation of γδ Cell Differentiation by αβ T Cell Progenitors

Bruno Silva-Santos,* Daniel J. Pennington,*† Adrian C. Hayday†

The thymus gives rise to two T cell lineages, αβ and γδ, that are thought to develop independently of one another. Hence, double positive (DP) thymocytes expressing CD4 and CD8 coreceptors are usually viewed simply as progenitors of $CD4^+$ and $CD8^+$ αβ T cells. Instead we report that DP cells regulate the differentiation of early thymocyte progenitors and γδ cells, by a mechanism dependent on the transcription factor RORγt, and the lymphotoxin (LT) β receptor (LTβR). This finding provokes a revised view of the thymus, in which lymphoid tissue induction-type processes coordinate the developmental and functional integration of the two T cell lineages.

Cell-mediated immunity involves T cell receptor (TCR) $\alpha\beta^+$ cells, which recognize antigenic peptides presented by major histocompatibility complex (MHC) proteins, and unconventional, non–MHC-restricted T cells, of which $TCR\gamma\delta^+$ cells are the prototype. There is increasing evidence that αβ and γδ T cells are functionally integrated, but their relatedness and how they may "cross-talk" are incompletely understood (*1*). To clarify this situation, we identified a "γδ-biased" gene profile (*2*). Unexpectedly, full expression of this profile by $TCR\gamma\delta^+$ thymocytes depended in trans on $CD4^+CD8^+$ (DP) cells, which are late-stage αβ T cell progenitors that form the most abundant thymocyte subset (Fig. 1A). Reflecting this situation, γδ cell function is altered in $TCR\beta^{-/-}$ mice that lack normal DPs (*2*).

We hypothesized that DP cells might exert their effects directly on maturing γδ T

Peter Gorer Department of Immunobiology, Guy's King's St. Thomas' Medical School, King's College, Guy's Hospital, London SE1 9RT, UK.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: daniel.pennington@kcl.ac.uk (D.J.P.), adrian.hayday@kcl.ac.uk (A.C.H.)

active against both nematodes and insects, requires the *bre* pathway for full activity against *C. elegans* (*13*, *14*). Taken together, our data and these studies suggest that both nematicidal and insecticidal three-domain Bt toxins use invertebrate glycolipids as host cell receptors and that the loss of glycolipid receptors represents an important mechanism for Bt toxin resistance. The ease with which glycolipids can be isolated and analyzed suggests that it will be feasible to monitor glycolipid-mediated resistance in field and laboratory populations of insects and nematodes.

Glycolipids, however, are not the only class of Bt toxin host cell receptors. For insects, high-affinity protein receptors, such as insect cadherins and aminopeptidases, have been shown to play functional roles as Cry1 receptors, although correlating protein receptor defects with binding defects has not always been simple. For example, in at least two cases, Cry1Ac has been found to bind specifically to membranes of cadherin receptor mutants that are Cry1Ac resistant (*26*, *27*), leading to the hypothesis that a multistep binding process involving multiple receptors is required for proper pore formation. For nematodes, our data suggest that although *bre*-dependent glycolipids are important for Cry14A function, there is likely another binding factor involved in Cry14A toxicity (*14*). We hypothesize that glycolipid and protein receptors may both play a role, sequentially or simultaneously, in positioning Bt toxins appropriately at the bilayer or in inserting toxins into the bilayer.

Mammalian cells do not bind three-domain Bt Cry toxins (*2*), and the results presented here provide a plausible molecular basis for the lack of toxicity of Cry toxins toward vertebrates. Vertebrates lack arthroseries glycolipids, which contain the conserved invertebrate-specific core tetrasaccharide GalNAcβ1-4GlcNAcβ1-3Manβ1-4Glc that is synthesized by the BRE pathway. Although the β-linked galactose important for Cry5B binding is not present in this core sequence, our unpublished data indicate that the intact receptor is, by three orders of magnitude, a better competitive inhibitor than β-methylgalactoside. Thus, higher-order structure is likely important for binding. We hypothesize that three-domain Bt Cry toxins evolved to at least partly recognize the core arthroseries tetrasaccharide and thus target the invertebrates, nematodes, and insects that synthesize these molecules.

The high degree of conservation between glycolipids present in *C. elegans* and in the human parasitic nematodes *Ascaris suum* and *Onchocerca volvolus*, which are phylogenetically divergent from *C. elegans*, suggests that most, if not all, nematodes will be susceptible to Cry5B toxin. All the nematodes we have tested to date are susceptible to Cry5B, and the one animal parasite we have tested has been shown to succumb to both Cry5B and Cry14A (*5*). Given these data and evidence that Cry5B and Cry14A toxin use an invertebrate-type glycolipid as their receptor, these Cry proteins hold great promise for one day safely targeting nematode pests of animals and plants.

References and Notes

1. R. A. de Maagd, A. Bravo, N. Crickmore, *Trends Genet.* **17**, 193 (2001).
2. F. S. Betz, B. G. Hammond, R. L. Fuchs, *Regul. Toxicol. Pharmacol.* **32**, 156 (2000).
3. M. Qaim, D. Zilberman, *Science* **299**, 900 (2003).
4. C. James, *ISAAA Briefs* **27**, 1 (2002).
5. J. Z. Wei *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 2760 (2003).
6. F. Gould, *Annu. Rev. Entomol.* **43**, 701 (1998).
7. L. J. Gahan, F. Gould, D. G. Heckel, *Science* **293**, 857 (2001).
8. S. Morin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5004 (2003).
9. R. Rajagopal, S. Sivakumar, N. Agrawal, P. Malhotra, R. K. Bhatnagar, *J. Biol. Chem.* **277**, 46849 (2002).
10. B. Oppert, K. J. Kramer, R. W. Beeman, D. Johnson, W. H. McGaughey, *J. Biol. Chem.* **272**, 23473 (1997).
11. D. G. Heckel, L. C. Gahan, F. Gould, A. Anderson, *J. Econ. Entomol.* **90**, 75 (1997).
12. J. L. Jurat-Fuentes, F. L. Gould, M. J. Adang, *Appl. Environ. Microbiol.* **68**, 5711 (2002).
13. J. S. Griffitts, J. L. Whitacre, D. E. Stevens, R. V. Aroian, *Science* **293**, 860 (2001).
14. J. S. Griffitts *et al.*, *J. Biol. Chem.* **278**, 45594 (2003).
15. L. D. Marroquin, D. Elyassnia, J. S. Griffitts, J. S. Feitelson, R. V. Aroian, *Genetics* **155**, 1693 (2000).
16. R. A. de Maagd, A. Bravo, C. Berry, N. Crickmore, H. E. Schnepf, *Annu. Rev. Genet.* **37**, 409 (2003).
17. J. S. Griffitts *et al.*, data not shown.
18. T. Yang, O. K. Baryshnikova, H. Mao, M. A. Holden, P. S. Cremer, *J. Am. Chem. Soc.* **125**, 4779 (2003).
19. B. E. Collins, J. C. Paulson, *Curr. Opin. Chem. Biol.* **8**, 617 (2004).
20. J. Mucha *et al.*, *Biochem. J.* **382**, 67 (2004).
21. S. Gerdt, R. D. Dennis, G. Borgonie, R. Schnabel, R. Geyer, *Eur. J. Biochem.* **266**, 952 (1999).
22. Z. S. Kawar, I. Van Die, R. D. Cummings, *J. Biol. Chem.* **277**, 34924 (2002).
23. S. F. Garczynki, M. J. Adang, in *Entomopathogenic Bacteria: From Laboratory to Field Application*, J.-F. Charles, A. Delecluse, C. Nielsen-LeRoux, Eds. (Kluwer, Netherlands, 2000), pp. 181–197.
24. N. S. Kumaraswami *et al.*, *Comp. Biochem. Physiol. B* **129**, 173 (2001).
25. R. D. Dennis, H. Wiegandt, D. Haustein, B. H. Knowles, D. J. Ellar, *Biomed. Chromatogr.* **1**, 31 (1986).
26. M. K. Lee, F. Rajamohan, F. Gould, D. H. Dean, *Appl. Environ. Microbiol.* **61**, 3836 (1995).
27. J. Gonzalez-Cabrera, B. Escriche, B. E. Tabashnik, J. Ferre, *Insect Biochem. Mol. Biol.* **33**, 929 (2003).
28. Materials and methods are available as supporting material on *Science* Online.

29. We thank R. Schnaar for technical advice on glycolipids, B. Hayes and N. Preece of the UCSD Glycotechnology Core for help with glycan analysis, T. Huxford for assistance with toxin purification, D. Huffman for comments, and members of the R.V.A. laboratory for stimulating discussions. This work was supported by NSF grant MCB-9983013 and grants from the Burroughs–Wellcome Foundation and the Beckman Foundation (to R.V.A.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/307/5711/922/DC1
Materials and Methods
SOM Text
Figs. S1 to S5
Tables S1 to S5
References

24 August 2004; accepted 19 November 2004
10.1126/science.1104444

Lymphotoxin-Mediated Regulation of γδ Cell Differentiation by αβ T Cell Progenitors

Bruno Silva-Santos,* Daniel J. Pennington,*† Adrian C. Hayday†

The thymus gives rise to two T cell lineages, αβ and γδ, that are thought to develop independently of one another. Hence, double positive (DP) thymocytes expressing CD4 and CD8 coreceptors are usually viewed simply as progenitors of $CD4^+$ and $CD8^+$ αβ T cells. Instead we report that DP cells regulate the differentiation of early thymocyte progenitors and γδ cells, by a mechanism dependent on the transcription factor RORγt, and the lymphotoxin (LT) β receptor (LTβR). This finding provokes a revised view of the thymus, in which lymphoid tissue induction-type processes coordinate the developmental and functional integration of the two T cell lineages.

Cell-mediated immunity involves T cell receptor (TCR) $\alpha\beta^+$ cells, which recognize antigenic peptides presented by major histocompatibility complex (MHC) proteins, and unconventional, non–MHC-restricted T cells, of which $TCR\gamma\delta^+$ cells are the prototype. There is increasing evidence that αβ and γδ T cells are functionally integrated, but their relatedness and how they may "cross-talk" are incompletely understood (*1*). To clarify this situation, we identified a "γδ-biased" gene profile (*2*). Unexpectedly, full expression of this profile by $TCR\gamma\delta^+$ thymocytes depended in trans on $CD4^+CD8^+$ (DP) cells, which are late-stage αβ T cell progenitors that form the most abundant thymocyte subset (Fig. 1A). Reflecting this situation, γδ cell function is altered in $TCR\beta^{-/-}$ mice that lack normal DPs (*2*).

We hypothesized that DP cells might exert their effects directly on maturing γδ T

Peter Gorer Department of Immunobiology, Guy's King's St. Thomas' Medical School, King's College, Guy's Hospital, London SE1 9RT, UK.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: daniel.pennington@kcl.ac.uk (D.J.P.), adrian.hayday@kcl.ac.uk (A.C.H.)

cells, just as they can affect natural killer T (NKT) cell selection (*3*), or on early γδ cell progenitors. To test this hypothesis, we tracked the onset of expression of a subset of γδ-biased genes that is underexpressed in TCRβ$^{-/-}$ mice lacking normal DP cells (*2*). The subset included four transcription factors (ICER, Nor1, Nurr1, and Nur77) and two regulators of heterotrimeric GTP-binding protein (G protein) signaling (Rgs1 and Rgs2). We found that the least differentiated, CD4$^-$CD8$^-$ [double negative (DN)] thymocytes, which progress through four stages (DN1 to DN4) before becoming either DPs or TCRγδ$^+$ thymocytes (Fig. 1A), already expressed all the profiled genes (Fig. 1B). Moreover, in mice in which *lacz* is expressed from the ICER promoter (*4*), intracellular β-galactosidase (*5*) was detected in essentially all DN2 cells (Fig. 1C). Hence, the expression of ICER (and possibly other γδ-biased genes) is a feature of DN2 cells in general and not just a minor subset.

DN-to-DP maturation requires a pre-TCR that includes TCRβ [encoded following recombinase activating gene (RAG)–mediated TCRβ gene rearrangement], a pre-Tα (pTα) chain, and p56^{Lck} kinase (*6*–*9*). Hence, RAG2$^{-/-}$ (*10*), TCRβ$^{-/-}$ (*11*), pTα$^{-/-}$ (*9*), and p56$^{Lck-/-}$ mice (*12*) all provide opportunities to examine γδ-biased gene expression in DN2 and DN3 progenitors that develop in the absence of normal DPs. Each mutant showed reductions in γδ-biased gene expression, ranging from ~4-fold for Nur77 to >32-fold for ICER, as judged by multiple, independent reverse transcriptase–polymerase chain reaction (RT-PCR) titrations and real-time PCR (*5*) (Fig. 1, D and E; fig. S1). Because DN2 thymocytes do not express pre-TCR complexes (*13*, *14*), the influence of the pre-TCR on their expression of γδ-biased genes must reflect a trans effect imposed on DN2 cells presumably via DP cells. Consistent with this hypothesis, TCRα$^{-/-}$ (*11*) and TCRδ$^{-/-}$ (*15*) mice, which retain full DP compartments, displayed normal ICER expression (Fig. 1E). The reduced expression of γδ-biased genes in thymocytes from DP-deficient mice was associated with alterations in peripheral γδ cell function, because short-term stimulation of splenic γδ cells from either TCRβ$^{-/-}$ mice or pTα$^{-/-}$ mice provoked markedly reduced interferon-γ (IFN-γ), tumor necrosis factor α (TNF-α), and interleukin-2 (IL-2) production (Fig. 1F), extending previous long-term stimulation assays of γδ cells from TCRβ$^{-/-}$ mice (*2*).

Embryonic lymphoid tissue inducer (LTi) cells are a recently discovered component of lymphoid development. They produce the heteromeric cytokine lymphotoxin (LT) (e.g., $LT\alpha_1LT\beta_2$) that stimulates LT-receptor (LTβR)–expressing mesenchymal cells to produce factors that organize lymphoid cells into lymph nodes (LNs) and Peyer's patches (PPs) (*16*–*18*). LTi cells express the transcription factor RORγt (*19*), and RORγt$^{-/-}$ (*20*), LTα$^{-/-}$ (*16*), LTβ$^{-/-}$ (*21*), and LTβR$^{-/-}$ (*18*) mice lack normal LNs and PPs. DP thymocytes are one of only a few cell types to express RORγt postnatally (*19*), suggesting that they may have LTi activity.

Consistent with this possibility, mRNAs for RORγt, LTα, LTβ, and for LIGHT (another LTβR ligand) are up-regulated by pre-TCR–associated signaling in vivo (fig. S2) and are expressed by DP thymocytes (Fig. 2A). Conversely, rare DP cells that form in TCRβ$^{-/-}$ mice independently of the pre-TCR (*22*), and that do not support normal γδ differentiation (*2*), show greatly reduced RORγt, LTβ, and LIGHT expression (Fig. 2A). Although LTα expression was unaffected, it alone is insufficient to activate LTβR (*23*). The hypothesis that LTi-like activities condition γδ cell differentiation received further support from studies of RORγt$^{-/-}$ mice. Whereas the RORγt$^{-/-}$ thymus is disorganized, TCRγδ$^+$ thymocytes are normal in number, viability, and gross morphology (fig. S3). Nonetheless, they largely phenocopy TCRβ$^{-/-}$ mice in show-

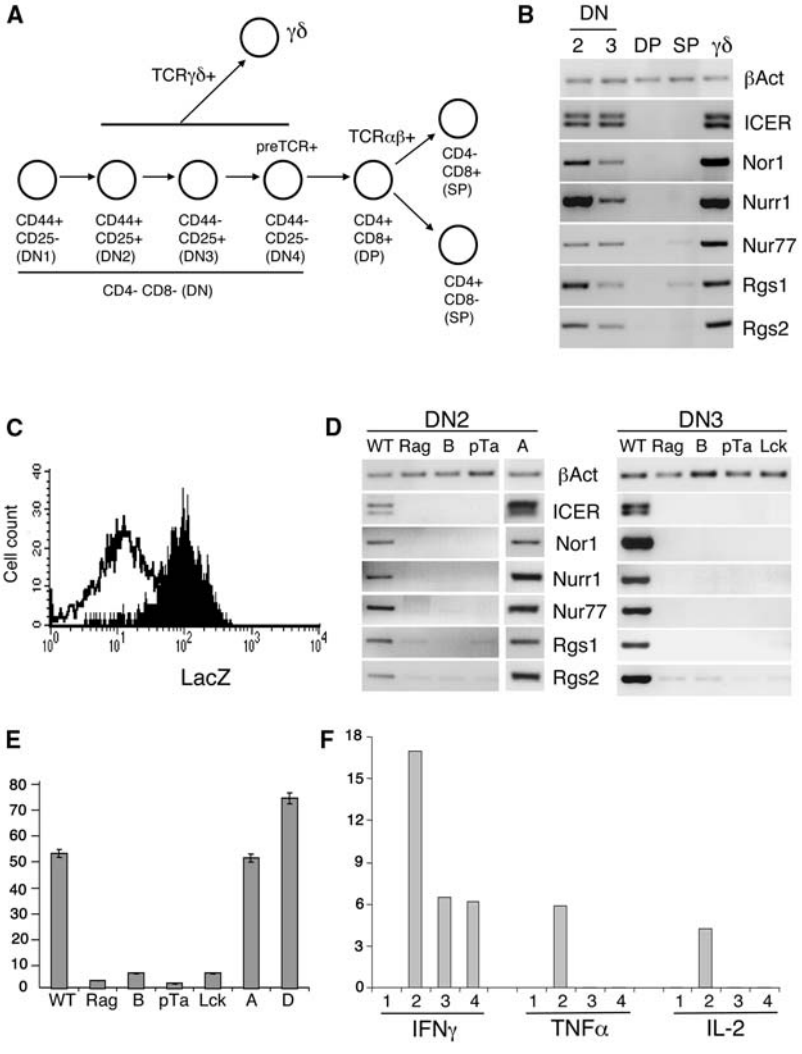


Fig. 1. Mice lacking β-selected DP thymocytes show impaired γδ cell responses to activation and fail to express a γδ-biased gene profile in mature γδ thymocytes and their progenitors. (**A**) Developmental stages of mouse thymocytes. (**B**) RT-PCR for γδ-biased genes in thymocyte subsets of wild-type (WT) mice. DN denotes CD4$^-$CD8$^-$ TCRγδ$^-$ NK1.1$^-$ B220$^-$ cells, and SP denotes CD4$^+$ CD8$^-$ thymocytes. (**C**) Staining for β-galactosidase expression in DN2 thymocytes (shaded) of mice heterozygous for the insertion of *lacz* downstream of the ICER promoter [negative control (white)]. (**D**) RT-PCR for the γδ-biased gene panel in DN2 and DN3 thymocytes of WT, Rag2$^{-/-}$ (Rag), TCRβ$^{-/-}$ (lane B), pTα$^{-/-}$ (pTa), TCRα$^{-/-}$ (lane A), and p56Lck$^{-/-}$ (Lck) mice. (**E**) Quantitative real-time PCR (presented as ICER/glyceraldehyde-3-phosphate dehydrogenase ratio) for ICER expression in DN3 thymocytes of mice described for (D) and in TCRδ$^{-/-}$ mice (bar D). Error bars represent standard errors of the means of triplicate samples. (**F**) Percentage of γδ splenocytes expressing intracellular IFN-γ, TNF-α, or IL-2 proteins after 18 hours in culture. 1, resting WT cells; 2 to 4, activated with CD3ε monoclonal antibody (mAb) (2, WT; 3, TCRβ$^{-/-}$; 4, pTα$^{-/-}$ splenocytes). Values are increase above nonactivated control (1) and are representative of pooled cells from 6 to 10 mice.

ing reduced expression of the profiled gene set (Fig. 2B), as well as of granzyme B (GzmB), a gene associated with cytolytic function that is also γδ-biased and underexpressed in TCRβ$^{-/-}$ mice (*2*).

To test whether the profiled gene set is regulated by LT, we cultured late fetal thymic lobes from TCRβ$^{-/-}$ mice (*5*, *24*) together with an antibody against LTβR (anti-LTβR) agonist (*25*). The antibody did not affect the numbers of TCRγδ^{+} cells that developed, but it did rescue the cells' expression of γδ-biased genes (Fig. 3A), including Dap12, a γδ-biased signaling molecule that is also underexpressed in TCRβ$^{-/-}$ mice (*2*). Control gene *Egr1*, which is expressed comparably by TCRγδ^{+} thymocytes from wild-type (WT) or TCRβ$^{-/-}$ mice, was unaffected, whereas CD25, which is expressed slightly less in WT mice, was slightly down-regulated by the LTβR agonist (Fig. 3A). In sum, LTβR could restore a normal γδ phenotype, complementing dysregulation in the DP compartment.

Consistent with this result, the profiled genes were down-regulated in WT TCRγδ^{+} thymocytes developing in the presence of a soluble LTβR-Fc receptor fusion protein (*26*) that competes for ligand binding to LTβR (Fig. 3B). Although the effects on Rgs2 and Nur77 were less overt, this may be because LT blocking was incomplete and because Rgs2 and Nur77 may be less dependent on LT for their induction. To study the effects of blocking LT genetically, we examined LTβR$^{-/-}$ (*18*), LTβ$^{-/-}$ (*21*), and LTα$^{-/-}$ mice (*16*). In each case, the numbers, morphology, and viability of recoverable TCRγδ^{+} thymocytes were normal (fig. S3), but again the cells' expression of γδ-biased genes was reduced (Fig. 3C and fig. S4). Rgs2 was least affected, mirroring the fusion protein blocking result (Fig. 3B).

An effect of LT/LTβR signaling on γδ cell function was evident in that splenic γδ cells from either LTβR$^{-/-}$ or LTα$^{-/-}$ mice failed to show normal activation-induced up-regulation of RNAs for GzmA, Dap12, and cyclin D3 (which is normally associated with T cell proliferation) (Fig. 3D). We next examined lethally irradiated WT or LTα$^{-/-}$ mice that had been reconstituted with either WT or LTα$^{-/-}$ bone marrow (*5*). One month after engraftment, splenic γδ cells were harvested, stimulated via the TCR for 18 hours, and examined before cell division. About 10% of γδ cells from the WT→WT chimeras displayed activation-induced IFN-γ and IL-2 production, with similar results in WT→LTα$^{-/-}$ chimeras. Conversely, negligible numbers of cells in either LTα$^{-/-}$→WT or LTα$^{-/-}$→LTα$^{-/-}$ chimeras induced IFN-γ and IL-2 (Fig. 3E). Thus, bone marrow–derived LTα determines the normal differentiation of TCRγδ^{+} thymocytes and the consequent γδ cell phenotype.

In lymphoid organogenesis, LT produced by LTi cells provokes LTβR^{+} stromal cells to express molecules, including ICAM1 (intercellular adhesion molecule 1), that locally organize lymphoid cells (*17*). Similarly, LT produced by thymocytes can condition medullary epithelial cell differentiation (*27*). Consistent with this observation, the LTβR agonist, which could rescue γδ-biased gene expression in TCRβ$^{-/-}$ thymic organ cultures, also up-regulated ICAM1 mRNA and surface expression on EpCAM^{+} (epithelial cell adhesion molecule) stromal cells (Fig. 4, A and B). This result notwithstanding, DN and DP cells reside in the thymic cortex, where epithelial cells are reportedly unaffected by LT (*27*), thus raising the possibility that DP cells might act directly on cortical thymocytes.

In support of this hypothesis, we found that DN2, DN3, and TCRγδ^{+} thymocytes from WT or TCRβ$^{-/-}$ mice expressed LTβR mRNA (Fig. 4C and fig. S5). Moreover, we found that DN2/DN3 and TCRγδ^{+} thymocytes from TCRβ$^{-/-}$ mice were able to up-regulate γδ-biased genes following coculture with DP cells from WT or TCRδ$^{-/-}$ mice (Fig. 4D and fig. S6). After coculture with equivalent numbers of atypical DP cells from TCRβ$^{-/-}$ mice that do not express normal levels of heteromeric LTβR ligands, the recovery of viable TCRγδ^{+} thymocytes was comparable (fig. S7), but the cells did not up-regulate the γδ-biased genes unless an agonist anti-LTβR was added (Fig. 4D). This result confirmed that the primary functional defect of DP cells from TCRβ$^{-/-}$ mice is reduced expression of LTβR ligands. Nevertheless, the LTβR agonist

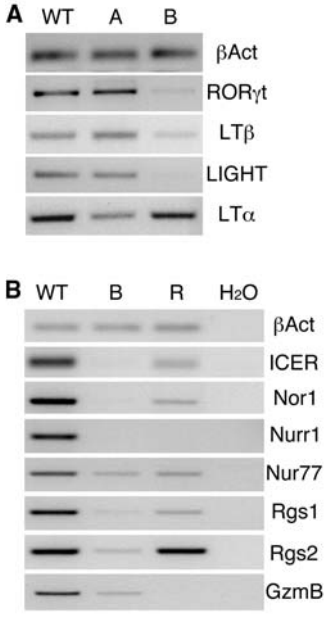


Fig. 2. RORγt is expressed in β-selected DP thymocytes and is required for the expression of a γδ-biased gene profile in γδ thymocytes. (**A**) RT-PCR for expression of RORγt, LTα, LTβ, and LIGHT in DP thymocytes of WT, TCRα$^{-/-}$ (lane A), and TCRβ$^{-/-}$ (lane B) mice. (**B**) Expression of γδ-biased genes in γδ thymocytes of RORγt$^{-/-}$ mice (lane R), compared to WT and TCRβ$^{-/-}$ mice (lane B).

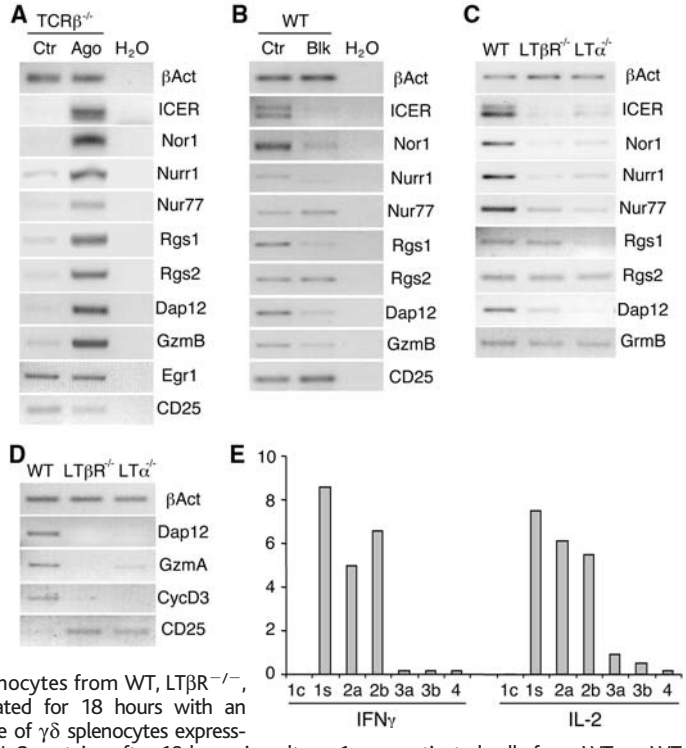


Fig. 3. The lymphotoxin pathway is required for normal gene expression and responsiveness of γδ cells. (**A**) RT-PCR for expression of γδ-biased genes in γδ thymocytes from TCRβ$^{-/-}$ fetal thymic lobes cultured in fetal thymic organ culture (FTOC) for 7 days with anti-LTβR agonist (Ago) or with isotype control (Ctr). (**B**) Expression of γδ-biased genes in γδ thymocytes from WT fetal thymic lobes cultured in FTOC for 7 days in the presence of an LTβR:Fc fusion protein (Blk) or Fc fusion protein control (Ctr). (**C**) Expression of γδ-biased genes in γδ thymocytes ex vivo from WT, LTβR$^{-/-}$, and LTα$^{-/-}$ mice. (**D**) Expression of activation-induced genes in γδ splenocytes from WT, LTβR$^{-/-}$, and LTα$^{-/-}$ mice, activated for 18 hours with an CD3ε mAb. (**E**) Percentage of γδ splenocytes expressing intracellular IFN-γ or IL-2 proteins after 18 hours in culture. 1c, nonactivated cells from WT → WT bone marrow chimerae; 1s to 4, cells activated with CD3ε mAb, isolated from the following bone marrow chimerae: 1s, WT → WT; 2a and 2b, WT → LTα$^{-/-}$; 3a and 3b, LTα$^{-/-}$ → WT; 4, LTα$^{-/-}$ → LTα$^{-/-}$. Values are increase above nonactivated control (1c) and are representative of four to eight mice.

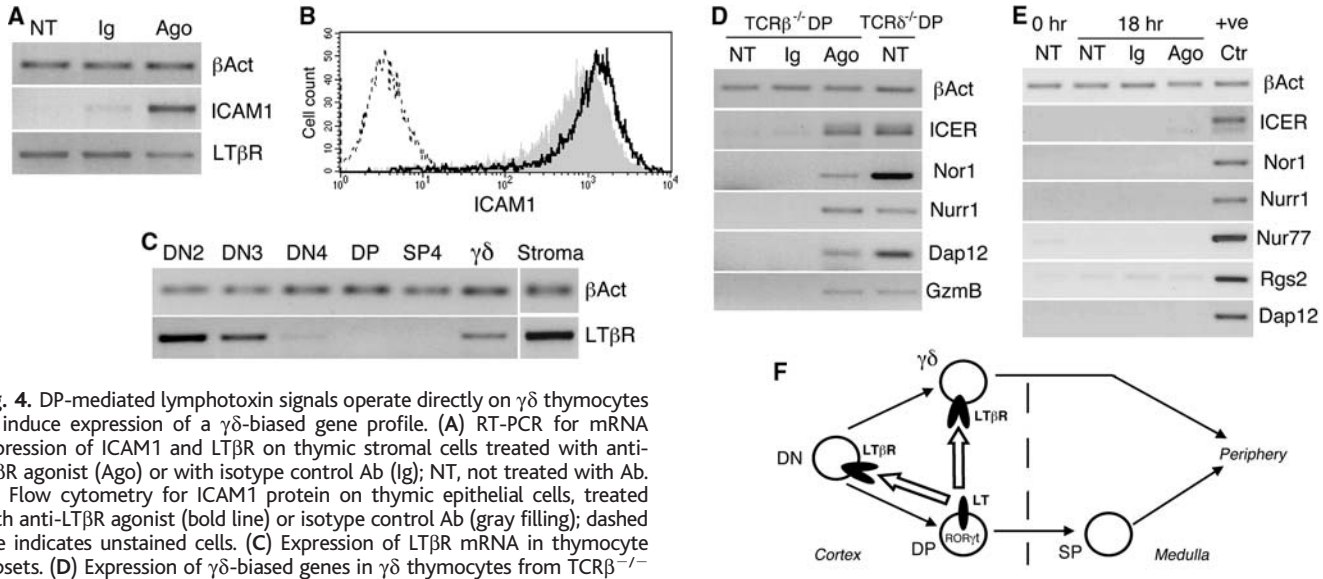


Fig. 4. DP-mediated lymphotoxin signals operate directly on γδ thymocytes to induce expression of a γδ-biased gene profile. (**A**) RT-PCR for mRNA expression of ICAM1 and LTβR on thymic stromal cells treated with anti-LTβR agonist (Ago) or with isotype control Ab (Ig); NT, not treated with Ab. (**B**) Flow cytometry for ICAM1 protein on thymic epithelial cells, treated with anti-LTβR agonist (bold line) or isotype control Ab (gray filling); dashed line indicates unstained cells. (**C**) Expression of LTβR mRNA in thymocyte subsets. (**D**) Expression of γδ-biased genes in γδ thymocytes from TCRβ$^{-/-}$ mice cocultured for 18 hours with DP cells isolated from either TCRβ$^{-/-}$ or control TCRδ$^{-/-}$ mice in the presence of anti-LTβR agonist (Ago) or isotype control Ab (Ig); NT, not treated with Ab. (**E**) Expression of γδ-biased genes in γδ thymocytes from TCRβ$^{-/-}$ mice incubated alone for 18 hours with anti-LTβR agonist (Ago) or isotype control Ab (Ig), NT; not treated with Ab. (**F**) Modified scheme for murine thymocyte development.

alone failed to induce the profiled genes, revealing that other factors contribute to the trans effect of DP cells on developing thymocytes (Fig. 4E).

Collectively, these data paint a picture of the thymus in which DP cells, until now viewed simply as T cell progenitors, have LTi-like capacity to regulate early thymocyte progenitors (Fig. 4F). Reflecting this situation, peripheral γδ cell responses are altered in mice lacking either normal DPs or LTi capacity. Although stroma may mediate thymocyte cross-talk in vivo, we demonstrate direct LT-mediated DP-DN/γδ cell interactions, possibly related to which DN progenitors first enter the thymus in a region rich in DP cells (*28*).

The feedback regulation of DN progenitors by DP cells is reminiscent of genetic epistasis and cautions against the view that thymocytes from RAG$^{-/-}$, TCRβ$^{-/-}$, and pTα$^{-/-}$ mice are normal up until the DN3/DN4 stage at which the RAG, TCRβ, and pTα genes act cell-autonomously. Furthermore, the finding that ICER is expressed by most DN2 cells (Fig. 1C) implies that γδ-biased gene expression may not be limited to a small subset of putative precommitted γδ cell progenitors but may be a general property of early thymocytes that is subsequently down-regulated during conventional αβ T cell development. Indeed, as pre-TCR^{lo} DN4 cells mature into pre-TCR^{hi} DN4 cells that more readily mature into DP cells (*13*), their expression of γδ-biased genes decreases (fig. S8).

One can only speculate as to why DP cells regulate the development of other thymocytes. As the fetal thymus matures, it shifts from giving rise exclusively to γδ cells to yielding primarily conventional αβ T cells. Thus, DP cell numbers expand dramatically, potentially disrupting existing stromal-thymocyte interactions, for which the LTi-like activities of DP cells may compensate. Indeed, murine DP cells only express RORγt postnatally (*19*). Alternatively, LT-mediated developmental interdependence may optimize the subsequent functional integration of conventional and unconventional T cells. Indeed, the DP cells affect some genes more than others (Fig. 3 and fig. S1), and TGF-β, lymphotactin, and GzmC are expressed comparably by γδ cells from TCRβ$^{-/-}$ and WT mice (*14*). Likewise, DP cells may also condition the differentiation of unconventional αβ T cells that both express γδ-biased genes and regulate conventional αβ T cells (*2*, *29*).

Whereas most DP cells express RORγt, cell-bound LT is difficult to detect on them, and it may be only a subset that has LTi-like activity. Moreover, other thymocytes, including SP cells, may be a further source of LT, exerting effects in the thymic medulla (*27*). Nonetheless, our finding that the effects of DP cells are not solely attributable to LT (Fig. 4E) may explain why DP cells uniquely condition the expression of the γδ-biased genes studied here. Finally, LTi-like regulation of T cell differentiation may not be limited to DP thymocytes but might also be provided by other LTi cells, such as those in intestinal cryptopatches that have been implicated in unconventional T cell development (*30*).

References and Notes

1. A. Hayday, R. Tigelaar, *Nat. Rev. Immunol.* **3**, 233 (2003).
2. D. J. Pennington *et al.*, *Nat. Immunol.* **4**, 991 (2003).
3. A. Bendelac, *J. Exp. Med.* **182**, 2091 (1995).
4. J. A. Blendy, K. H. Kaestner, G. F. Weinbauer, E. Nieschlag, G. Schutz, *Nature* **380**, 162 (1996).
5. Materials and methods are available as supporting material on *Science* Online.
6. H. von Boehmer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 9729 (1988).
7. J. D. Marth *et al.*, *EMBO J.* **6**, 2727 (1987).
8. C. A. Mallick, E. C. Dudley, J. L. Viney, M. J. Owen, A. C. Hayday, *Cell* **73**, 513 (1993).
9. H. J. Fehling, A. Krotkova, C. Saint-Ruf, H. von Boehmer, *Nature* **375**, 795 (1995).
10. P. Mombaerts *et al.*, *Cell* **68**, 869 (1992).
11. P. Mombaerts *et al.*, *Nature* **360**, 225 (1992).
12. T. J. Molina *et al.*, *Nature* **357**, 161 (1992).
13. L. Bruno, A. Scheffold, A. Radbruch, M. J. Owen, *Curr. Biol.* **9**, 559 (1999).
14. B. Silva-Santos, D. J. Pennington, A. C. Hayday, data not shown.
15. S. Itohara *et al.*, *Cell* **72**, 337 (1993).
16. P. De Togni *et al.*, *Science* **264**, 703 (1994).
17. C. A. Cuff *et al.*, *J. Immunol.* **161**, 6853 (1998).
18. A. Futterer, K. Mink, A. Luz, M. H. Kosco-Vilbois, K. Pfeffer, *Immunity* **9**, 59 (1998).
19. G. Eberl *et al.*, *Nat. Immunol.* **5**, 64 (2004).
20. Z. Sun *et al.*, *Science* **288**, 2369 (2000).
21. M. B. Alimzhanov *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9302 (1997).
22. L. Passoni *et al.*, *Immunity* **7**, 83 (1997).
23. J. L. Gommerman, J. L. Browning, *Nat. Rev. Immunol.* **3**, 642 (2003).
24. E. J. Jenkinson, G. Anderson, J. J. Owen, *J. Exp. Med.* **176**, 845 (1992).
25. E. Dejardin *et al.*, *Immunity* **17**, 525 (2002).
26. I. A. Rooney *et al.*, *J. Biol. Chem.* **275**, 14307 (2000).
27. T. Boehm, S. Scheu, K. Pfeffer, C. C. Bleul, *J. Exp. Med.* **198**, 757 (2003).
28. E. F. Lind, S. E. Prockop, H. E. Porritt, H. T. Petrie, *J. Exp. Med.* **194**, 127 (2001).
29. P. Poussier, T. Ning, D. Banerjee, M. Julius, *J. Exp. Med.* **195**, 1491 (2002).
30. F. Lambolez *et al.*, *J. Exp. Med.* **195**, 437 (2002).

31. We thank A. Aguzzi, D. Littman, A. Smith, G. Schutz, M. Heikenwälder, T. Junt, A. McPherson, E. Niederer, F. Santori, M. Tsuji, S. Fen, W. Turnbull, J. Pang, E. Wise, A. Barnes, N. Dar, G. Anderson, D. Davies, and G. Warnes. This work was supported by the Wellcome Trust (B.S.-S., D.J.P, and A.C.H.) and the Dunhill Medical Trust (D.J.P.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/1103978/DC1
Materials and Methods
Figs. S1 to S8
References

12 August 2004; accepted 24 November 2004
Published online 9 December 2004;
10.1126/science.1103978
Include this information when citing this paper.

Mammalian SAD Kinases Are Required for Neuronal Polarization

Masashi Kishi,[1]* Y. Albert Pan,[1,2]* Justin Gage Crump,[3]† Joshua R. Sanes[1,2]‡

Electrical activity in neurons is generally initiated in dendritic processes then propagated along axons to synapses, where it is passed to other neurons. Major structural features of neurons—their dendrites and axons—are thus related to their fundamental functions: the receipt and transmission of information. The acquisition of these distinct properties by dendrites and axons, called polarization, is a critical step in neuronal differentiation. We show here that SAD-A and SAD-B, mammalian orthologs of a kinase needed for presynaptic differentiation in *Caenorhabditis elegans*, are required for neuronal polarization. These kinases will provide entry points for unraveling signaling mechanisms that polarize neurons.

Most vertebrate neurons have two types of cytoplasmic processes: shorter, tapering dendrites, which receive synaptic input, and longer, slender axons, which form synapses on target cells. The molecular bases of these structural and physiological distinctions have been studied in detail, but the mechanisms that establish polarization are poorly understood (*1*–*4*). The SAD-1 kinase SAD–1 has been identified in a screen for *Caenorhabditis elegans* mutants in which synapses failed to form properly, but some aspects of the mutant phenotype raised the possibility that it might also regulate neuronal polarization (*5*). Consistent with this idea, SAD-1 contains a kinase domain related to that of PAR–1, a determinant of embryonic polarity (*6*, *7*).

As part of an effort to identify regulators of synaptogenesis in mammals (*8*), we searched public databases for orthologs of *SAD-1* and found two each in humans and mice and one in *Drosophila*; a previously described Ascidian ortholog (*9*) is more distantly related (fig. S1A). We isolated the two murine genes (*10*), which we call *sad-a* and *sad-b*. Predicted SAD-A and -B proteins are similar to each other and to *C. elegans* SAD-1 along most of their length (Fig. 1A).

Northern analysis showed that expression of both *sad-a* and *-b* is largely restricted to the nervous system (Fig. 1B), and in situ hybridization revealed that both mRNAs are broadly distributed within the brain and spinal cord of embryonic and postnatal animals (Fig. 1, C and D, and fig. S1). *Sad* expression was low in the cerebral ventricular zone, in which neural progenitors divide, but high in newly generated neurons in the preplate [at embryonic day (E) 12] and in the cortical plate (at E16) (Fig. 1, E and F, and fig. S1). Thus, *sad* genes are activated early in the program of neuronal differentiation.

To determine the localization of the SAD proteins, we generated SAD-A–specific and SAD-B–specific antibodies (*10*). Specificity of the antibodies was demonstrated by reactivity with the corresponding recombinant protein and by the absence of immunoreactivity in tissue from SAD-null mutants described below. Both proteins were present in the gray matter of the brain and spinal cord and were present at, but not confined to, synaptic sites (Fig. 1, G and H). Staining of cultured hippocampal and cortical neurons (*10*) showed that SAD proteins were present in both axonal and dendritic processes (Fig. 1I and fig. S2).

To assess the function of the SAD kinases, we generated null mutants of both genes (Fig. 2A). *Sada*$^{-/-}$ and *sadb*$^{-/-}$ mice were healthy and fertile, even though lack of one kinase had no apparent compensatory effect on the expression of the other (Fig. 2B). In contrast, double-mutant pups (*sada*$^{-/-}$; *sadb*$^{-/-}$, called AB$^{-/-}$) showed little spontaneous movement, were only weakly responsive to

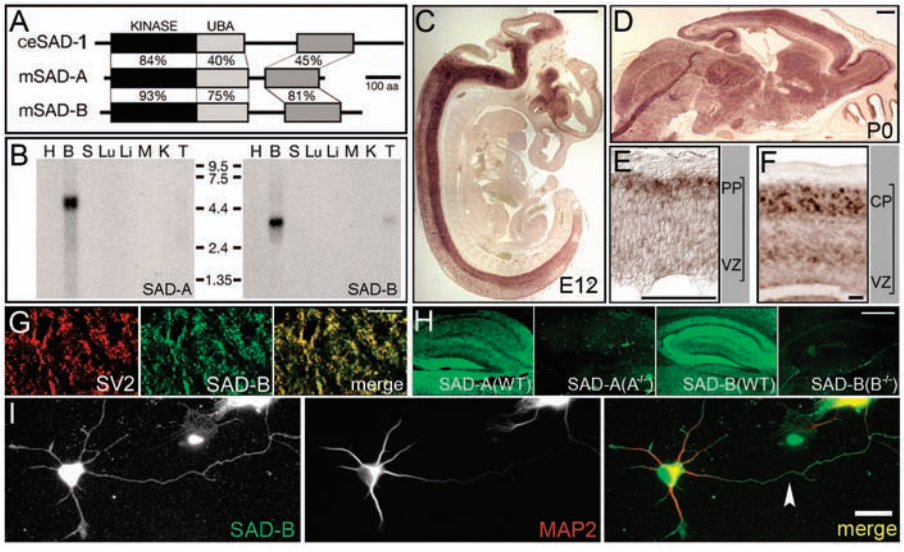


Fig. 1. Structure and expression of the mouse SAD kinases. (**A**) Homology of mouse (m) SAD-A and SAD-B (GenBank accession nos. AY533671 and AY533672) with each other and with *C. elegans* (Ce) SAD-1 includes the kinase domain, a region that contains a ubiquitin-associated (UBA) domain, and a C-terminal block of unique sequence. Numbers show percent amino acid identity. Scale bar, 100 amino acids. (**B**) Northern analysis of adult heart (H), brain (B), spleen (S), lung (Lu), liver (Li), skeletal muscle (M), kidney (K), and testis (T). SAD-A and -B are selectively expressed in the brain; low levels of SAD-B RNA are also detectable in the testis. (**C** and **D**) Expression of SAD-B by in situ hybridization of (C) E12 mouse embryo and (D) postnatal day 0 (P0) head. Both genes are broadly expressed in the brain and spinal cord. Scale bar, 500 μm. (**E** and **F**) Expression in the cortex is low in the proliferative ventricular zone (VZ) but high in (E) the preplate (PP) at E12 and (F) the cortical plate (CP) at E16, where newly generated neurons reside. Scale bar, 100 μm. (**G**) Section of adult spinal cord doubly stained with antibody to SAD-B and antibody to SV2, a synaptic vesicle protein. SAD-B is concentrated at synaptic sites in this region but is also present in nonsynaptic areas. Scale bar, 25 μm. (**H**) Sections of hippocampus from wild-type (WT), *sada*$^{-/-}$, and *sadb*$^{-/-}$ adults stained with antibodies specific for SAD-A or -B. Both proteins are broadly distributed in the neuropil. Specificity is shown by lack of staining in the cognate mutant. Scale bar, 500 μm. (**I**) Cultured hippocampal neurons stained with antibodies to SAD-B and the dendritic marker MAP2, showing SAD is present in both axon (arrowhead) and dendrites. Scale bar, 25 μm.

[1]Department of Anatomy and Neurobiology, Washington University Medical School, St. Louis, MO 63110, USA. [2]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA. [3]Department of Anatomy, University of California, San Francisco, CA 94143, USA.

*These authors contributed equally to this work.
†Present address: Institute of Neuroscience, University of Oregon, Eugene, OR 97403, USA.
‡To whom correspondence should be addressed. E-mail: sanesj@mcb.harvard.edu

tactile stimulation, and died within 2 hours of birth. Their immobility and poor responsiveness suggested a neural phenotype.

Examination of the $AB^{-/-}$ nervous system at E19 showed that the principal divisions of the brain, spinal cord, and peripheral nervous system had formed but that the forebrain was noticeably smaller in mutants than in littermate controls (fig. S3A). Although $AB^{-/-}$ cortex was abnormally thin (Fig. 2, C and D), it contained the main cortical cell types: neurons, astrocytes, and radial glia (fig. S3). The cells were organized into laminae, identifiable by cell density and specific markers (*11*, *12*). However, the subplate was poorly defined, and segregation of neuronal subtypes to sublayers within the cortical plate was disordered (Fig. 2D and fig. S3D) (*13*). It is therefore likely that SAD kinases are required for some early events in corticogenesis, but we focus here on the later aspect of neuronal differentiation.

To ask whether loss of SAD kinases affected neuronal morphology, we impregnated individual neurons with fluorescent dyes, using a ballistic transfer method (*14*). In controls, most labeled neurons were polarized, with a long, thick apical dendrite that extended toward the pia; multiple smaller dendrites arising from the soma; and a single, thin basal axon. In contrast, mutant neurons often had a starburst morphology or processes that ran diagonally or tangentially rather than radially, and axons were difficult to distinguish from dendrites (Fig. 2, E and F, and fig. S4).

To ask whether $AB^{-/-}$ cortical neurons extended long axons, we implanted crystals of 1,1′-dioctadecyl-3,3,3′,3′-tetramethylindocarbocyanine perchlorate (DiI) into the cortex (*15*). In controls, DiI-labeled axonal tracts descended from the cortex to the thalamus. In contrast, few if any labeled axons descended from $AB^{-/-}$ cortex (Fig. 2G). We also labeled sections with antibodies to transient axonal glycoprotein (TAG) 1, a selective marker of corticothalamic axons (*11*). TAG-1^{+} corticothalamic axons were evident in controls but absent from mutants (Fig. 2H). Although apoptosis was significantly increased in $AB^{-/-}$ cortex, the absence of the tract was not solely a consequence of neuronal loss, because in situ hybridization showed that TAG-1–expressing neurons were numerous in $AB^{-/-}$ mutants (fig. S3). We therefore favor the explanation that apoptosis is a consequence of axons failing to reach target areas and gain adequate trophic support (*16*).

To determine whether these morphological defects reflected defective neuronal polarity or were secondary consequences of other abnormalities, such as misplaced guidance cues, we grew neurons from $AB^{-/-}$ cortex in low-density cultures, then stained them with antibodies to microtubule-associated proteins (MAPs) that are selectively localized to either axons (dephospho-tau, stained with the Tau-1 antibody) or dendrites (MAP2) (*17*). We also analyzed cultures from the hippocampus (*17*, *18*), because polarity has been extensively studied in this system (*1*–*4*).

Neurons from mutant hippocampus (Fig. 3) and cortex (fig. S5) failed to form distinct axons and dendrites. Most control neurons had a single long, slender, branched axon that was rich in Tau-1 and poor in MAP2 and multiple shorter, thicker dendrites that were rich in MAP2 and poor in Tau-1 (Fig. 3A). In con-

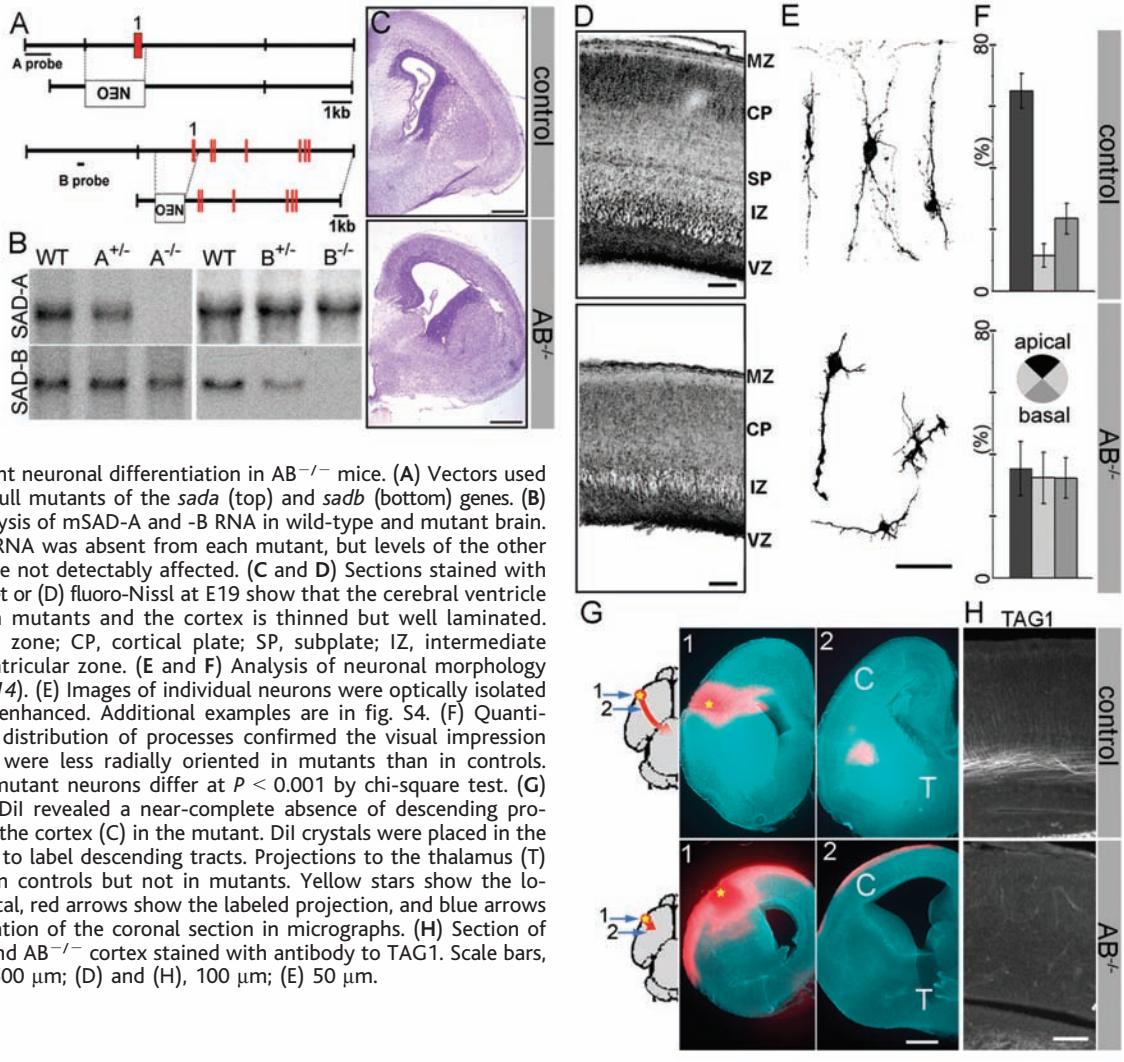


Fig. 2. Aberrant neuronal differentiation in $AB^{-/-}$ mice. (**A**) Vectors used to generate null mutants of the *sada* (top) and *sadb* (bottom) genes. (**B**) Northern analysis of mSAD-A and -B RNA in wild-type and mutant brain. The cognate RNA was absent from each mutant, but levels of the other SAD RNA were not detectably affected. (**C** and **D**) Sections stained with (C) cresyl violet or (D) fluoro-Nissl at E19 show that the cerebral ventricle is enlarged in mutants and the cortex is thinned but well laminated. MZ, marginal zone; CP, cortical plate; SP, subplate; IZ, intermediate zone; VZ, ventricular zone. (**E** and **F**) Analysis of neuronal morphology by diolistics (*14*). (E) Images of individual neurons were optically isolated and contrast-enhanced. Additional examples are in fig. S4. (F) Quantitation of the distribution of processes confirmed the visual impression that neurons were less radially oriented in mutants than in controls. Control and mutant neurons differ at $P < 0.001$ by chi-square test. (**G**) Tracing with DiI revealed a near-complete absence of descending projections from the cortex (C) in the mutant. DiI crystals were placed in the rostral cortex to label descending tracts. Projections to the thalamus (T) were visible in controls but not in mutants. Yellow stars show the location of crystal, red arrows show the labeled projection, and blue arrows show the location of the coronal section in micrographs. (**H**) Section of E19 control and $AB^{-/-}$ cortex stained with antibody to TAG1. Scale bars, (C) and (G), 500 μm; (D) and (H), 100 μm; (E) 50 μm.

trast, neurites of AB$^{-/-}$ neurons were relatively uniform in length and branching complexity, with values intermediate between those of axons and dendrites in control cultures (Fig. 3, D and E, and fig. S5). Moreover, most AB$^{-/-}$ neurites contained both Tau-1 and MAP2 (Fig. 3A), and levels of these markers did not differ greatly among neurites of a given neuron (Fig. 3, F and G, and fig. S5). Polarity defects did not result from growth defects, in that there was little difference between genotypes in the total number of primary neurites per neuron or the total length of neurite per neuron (Fig. 3, B and C, and fig. S5). Moreover, lack of polarization does not reflect a developmental delay, because control neurons were polarized by 3 days in vitro, whereas AB$^{-/-}$ neurons showed little polarization even after 8 days. These results indicate a cell-autonomous requirement of SAD kinases for neuronal polarization and suggest that the aberrant shape of AB$^{-/-}$ neurons in vivo results at least in part from this defect.

How do SAD kinases promote neuronal polarization? One possible mechanism is suggested by studies of PAR-1 and microtubule affinity-regulating kinases (MARKs), which have been implicated in the control of cellular polarity in worms, flies, and mammals (*6*, *7*, *19*). As noted above, SAD genes are related to *C. elegans* PAR-1 and its vertebrate orthologs, MARK1 to MARK4, in their kinase domains (50 to 52% amino acid identity to MARK1 to MARK4, compared to 93% identity between SAD-A and -B) but not in other domains. PAR-1 and MARK phosphorylate MAPs such as MAP2 and tau, thereby regulating their interactions with microtubules (*19*); local variations in microtubule organization are critical for neuronal polarization (*1*, *4*, *20*). We therefore asked whether SAD kinases also regulate phosphorylation of MAPs. We chose tau as an exemplar, because a specific site in this protein [serine 262 (S262)] is a substrate for PAR-1/MARK (*21*, *22*) and regulates binding of tau to microtubules (*23*). Overexpression of SAD-A in cultured neural (PC12) and nonneural (Chinese hamster ovary) cells increased phosphorylation of tau at S262 without affecting overall tau levels (Fig. 4A). A point mutation in the catalytic site (SAD-A^{K49A}), predicted to render SAD kinase inactive, abolished this effect (Fig. 4A). Thus, SAD kinase promotes phosphorylation of tau at S262 [tau-p(S262)] in vitro.

Next, we used immunostaining to examine the distribution of phosphorylated and dephosphorylated tau in cortex. Whereas dephospho-tau (recognized by Tau-1) is concentrated in axons of cultured neurons (*20*) (Fig. 3), tau-p(S262), like MAP2, is concentrated in dendrites (Fig. 4B). In the developing cortex, the upper cortical plate is rich in dendrites, whereas the intermediate zone is rich in axons but poor in dendrites (Fig. 4C). Consistent with this arrangement, tau-p(S262) and MAP-2 are concentrated in the upper cortical plate, whereas dephospho-tau is more abundant in the intermediate zone (Fig. 4D). We then asked whether tau phosphorylation is altered in AB$^{-/-}$ cortex, and we found that it is: Levels of tau-p(S262) are decreased and levels of dephospho-tau increased in the cortical plate, with no change in overall tau levels (Fig. 4, D and E). Together, these results suggest that SAD kinases act, at least in part, by locally regulating phosphorylation of MAPs, including tau, and that the consequent alterations in microtubule organization are critical for neuronal polarization (*20*). Because SAD is present in both axons and dendrites (Fig. 1I and fig. S2), its distinct effects on these two types of processes must reflect local regulation. Kinases such as MARK kinase/TAO1 and LKB1/PAR-4, which can activate MARK and SAD (*24*, *25*), are candidate local regulators. Other components of evolutionarily conserved polarization machinery, PAR-3 and PAR-6, which interact with PAR-1/MARK, may also be involved (*7*, *26*).

In summary, we have shown that the SAD kinases are required for forebrain neurons to acquire the polarity that endows axons and dendrites with distinct properties. SAD kinases seem to share some activities with PAR-1 and MARK, but their unique domains may

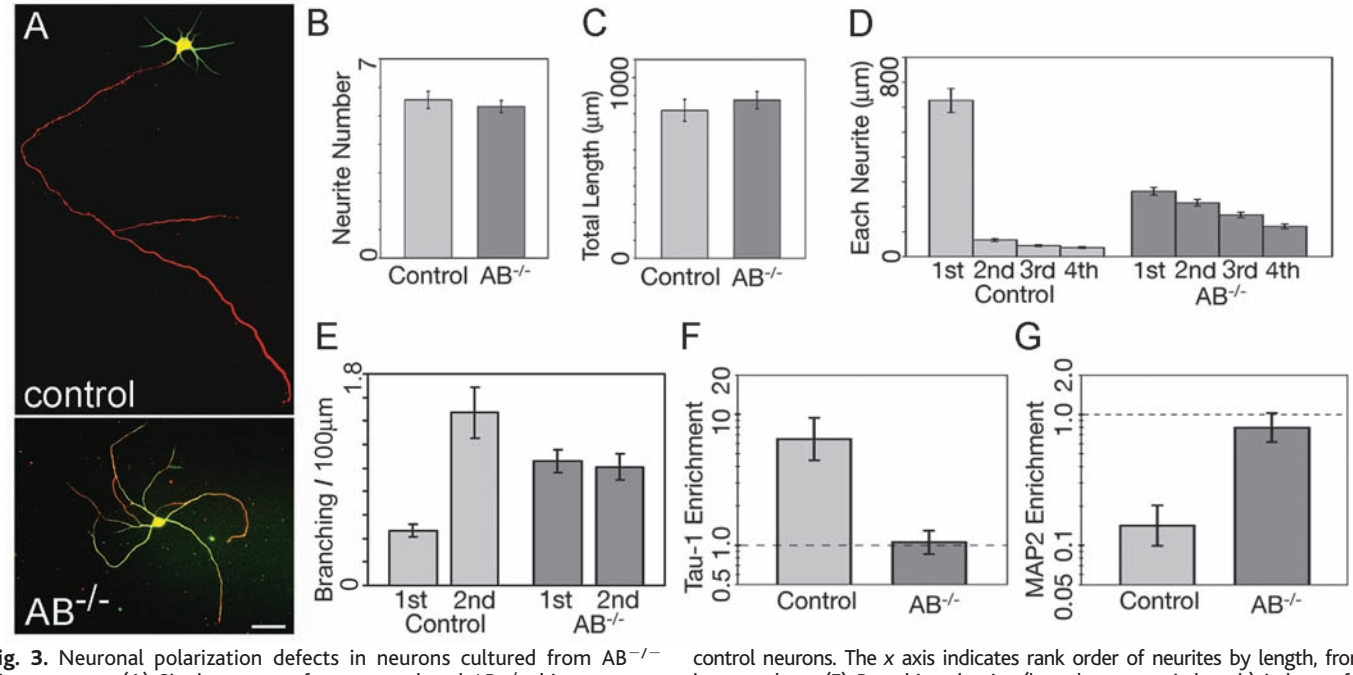


Fig. 3. Neuronal polarization defects in neurons cultured from AB$^{-/-}$ hippocampus. (**A**) Single neurons from control and AB$^{-/-}$ hippocampus cultured for 7 days, then stained for the axonal marker Tau-1 (red) and the dendritic marker MAP2 (green). Control neurons have a single, tau-rich axon and multiple, thicker, shorter, MAP2-rich dendrites. AB$^{-/-}$ neurons have multiple processes of similar length and caliber that contain both Tau-1 and MAP2. Scale bar, 50 μm. (**B** and **C**) Control and AB$^{-/-}$ neurons have equivalent numbers of neurites arising from the cell body and equivalent total neurite length. (**D**) Neurites of AB$^{-/-}$ neurons are similar in length to each other and midway in length between the axons and dendrites of control neurons. The *x* axis indicates rank order of neurites by length, from long to short. (**E**) Branching density (branches per unit length) is lower for axons (1st) than for the longest dendrite (2nd) in controls but is similar in AB$^{-/-}$ neurons. (**F** and **G**) Average intensity of Tau-1 or MAP2 staining in the longest neurite (the axon in control neurons) divided by that in all other neurites. Control axons are rich in Tau-1 and poor in MAP2; both antigens are more evenly distributed (the dashed line shows equivalence) in AB$^{-/-}$ neurons. Significance of differences between control and mutant neurons, (B) and (C) $P > 0.3$ by *t* test; (D) and (E) $P < 0.001$ by analysis of variance; and (F) and (G) $P < 0.001$ by permutation test. $n = 20$ neurons from two cultures.

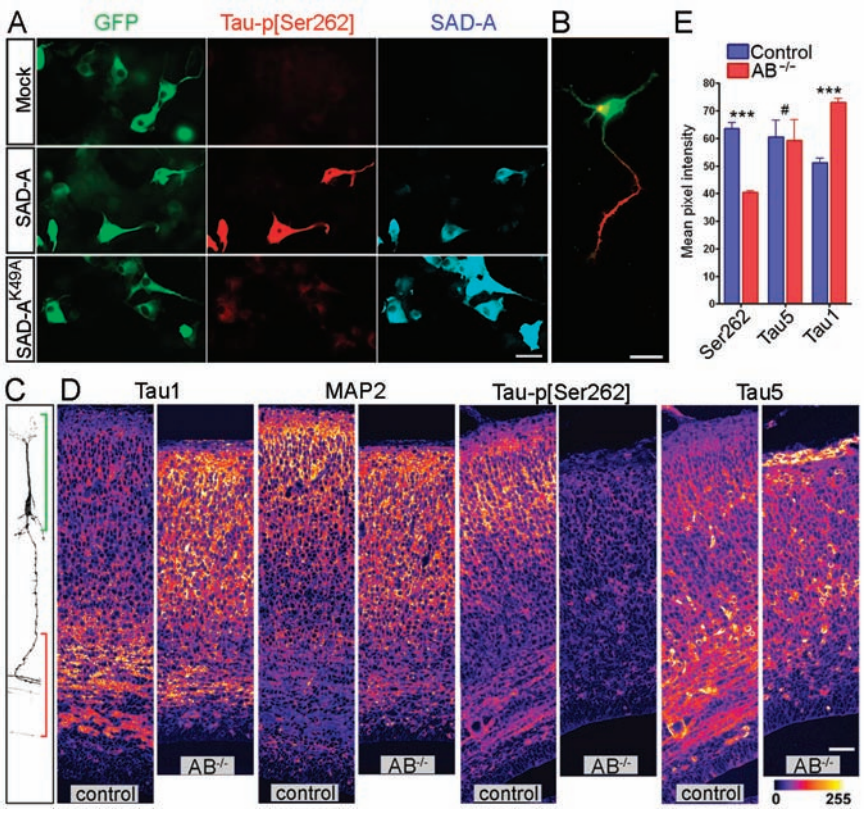


Fig. 4. SAD kinase regulates tau phosphorylation. (**A**) PC12 cells were transfected with a tau-GFP fusion alone (mock) or tau-GFP along with SAD-A or a kinase-defective SAD-A mutant (K49A). Three days later, cells were stained with antibody to SAD-A and antibody to tau-p(S262). SAD-A promotes tau phosphorylation at S262. Scale bar, 100 μm. (**B**) A cultured hippocampal neuron stained with antibody to tau-p(S262) (green) and antibody to dephospho-tau (Tau-1, red), showing concentration of the phosphoepitope in dendrites. Scale bar, 20 μm. (**C**) A single, DiI-labeled pyramidal neuron from a control to indicate the arrangement of neurons; the apical dendrite and axon are bracketed in green and red, respectively. (**D**) Sections of control and $AB^{-/-}$ cortex were labeled with antibodies to dephospho-tau (Tau-1), MAP2, tau-p(S262), or all tau isoforms (Tau5). Intensity is rendered in the color scale at the bottom right. In controls, MAP2 and tau-p(S262) were concentrated in the dendrite-rich upper cortical plate and Tau-1 in the axon-rich intermediate zone. In mutants, levels of dephospho-tau increased and tau-p(S262) decreased in the cortical plate; total tau levels did not change. Scale bar, 50 μm. (**E**) Quantitation of tau levels in the cortical plate. ***, $P < 0.001$; #, $P > 0.8$ by *t* test.

enable them to be locally regulated in neuron-specific ways. It remains to be determined whether mammalian SAD kinases, like *C. elegans* SAD-1 (*5*), are involved not only in polarity but also in synaptogenesis.

References and Notes

1. A. M. Craig, G. Banker, *Annu. Rev. Neurosci.* **17**, 267 (1994).
2. F. Bradke, C. G. Dotti, *Science* **283**, 1931 (1999).
3. M. A. Silverman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7051 (2001).
4. N. Arimura, C. Menager, Y. Fukata, K. Kaibuchi, *J. Neurobiol.* **58**, 34 (2004).
5. J. G. Crump, M. Zhen, Y. Jin, C. I. Bargmann, *Neuron* **29**, 115 (2001).
6. S. Guo, K. J. Kemphues, *Cell* **81**, 611 (1995).
7. A. Wodarz, *Nature Cell Biol.* **4**, E39 (2002).
8. J. R. Sanes, J. W. Lichtman, *Annu. Rev. Neurosci.* **22**, 389 (1999).
9. Y. Sasakura, M. Ogasawara, K. W. Makabe, *Mech. Dev.* **76**, 161 (1998).
10. Materials and methods are available as supporting material on *Science* Online.
11. R. F. Hevner *et al.*, *Dev. Neurosci.* **25**, 139 (2003).
12. O. Marin, J. L. Rubenstein, *Annu. Rev. Neurosci.* **26**, 441 (2003).
13. Y. A. Pan, M. Kishi, J. R. Sanes, in preparation.
14. W. B. Gan, J. Grutzendler, W. T. Wong, R. O. Wong, J. W. Lichtman, *Neuron* **27**, 219 (2000).
15. Z. Molnar, R. Adams, C. Blakemore, *J. Neurosci.* **18**, 5723 (1998).
16. B. Pettmann, C. E. Henderson, *Neuron* **20**, 633 (1998).
17. K. Goslin, H. Asmusse, G. Banker, in *Culturing Nerve Cells*, G. Banker, K. Goslin, Eds. (MIT Press, Cambridge, MA, ed. 2, 1998), pp. 339–370.
18. G. J. Brewer, J. R. Torricelli, E. K. Evege, P. J. Price, *J. Neurosci. Res.* **35**, 567 (1993).
19. G. Drewes, A. Ebneth, U. Preuss, E. M. Mandelkow, E. Mandelkow, *Cell* **89**, 297 (1997).
20. J. W. Mandell, G. A. Banker, *J. Neurosci.* **16**, 5727 (1996).
21. J. Biernat *et al.*, *Mol. Biol. Cell* **13**, 4013 (2002).
22. I. Nishimura, Y. Yang, B. Lu, *Cell* **116**, 671 (2004).
23. J. Biernat, N. Gustke, G. Drewes, E. M. Mandelkow, E. Mandelkow, *Neuron* **11**, 153 (1993).
24. T. Timm *et al.*, *EMBO J.* **22**, 5090 (2003).
25. J. Spicer, A. Ashworth, *Curr. Biol.* **14**, R383 (2004).
26. S. H. Shi, L. Y. Jan, Y. N. Jan, *Cell* **112**, 63 (2003).

27. We thank C. Bargmann for sharing data and helpful comments on the manuscript; R. Burgess, J. Cunningham, I.-J. Kim, T. Kummer, and R. Lewis for advice and assistance; G. Feng for a cDNA library; R. Hevner for antibodies; K. Kosik for green fluorescent protein (GFP)–tau cDNA; and the Mouse Genetics Core at Washington University for blastocyst injections. Supported by grants from NIH (J.R.S.) and an NIH National Research Service Award predoctoral fellowship (Y.A.P).

Supporting Online Material

www.sciencemag.org/cgi/content/full/307/5711/929/DC1
Materials and Methods
Figs. S1 to S5
References and Notes

11 November 2004; accepted 22 December 2004
10.1126/science.1107403

Methylation as a Crucial Step in Plant microRNA Biogenesis

Bin Yu,* Zhiyong Yang,* Junjie Li,† Svetlana Minakhina, Maocheng Yang, Richard W. Padgett, Ruth Steward, Xuemei Chen‡

Methylation on the base or the ribose is prevalent in eukaryotic ribosomal RNAs (rRNAs) and is thought to be crucial for ribosome biogenesis and function. Artificially introduced 2′-*O*-methyl groups in small interfering RNAs (siRNAs) can stabilize siRNAs in serum without affecting their activities in RNA interference in mammalian cells. Here, we show that plant microRNAs (miRNAs) have a naturally occurring methyl group on the ribose of the last nucleotide. Whereas methylation of rRNAs depends on guide RNAs, the methyltransferase protein HEN1 is sufficient to methylate miRNA/miRNA* duplexes. Our studies uncover a new and crucial step in plant miRNA biogenesis and have profound implications in the function of miRNAs.

MicroRNAs are 20- to 24-nucleotide (nt) RNAs that serve as sequence-specific regulators of gene expression in diverse eukaryotic organisms. miRNA/miRNA* duplexes, products of Dicer processing of pre-miRNAs, are structurally similar to small interfering RNAs

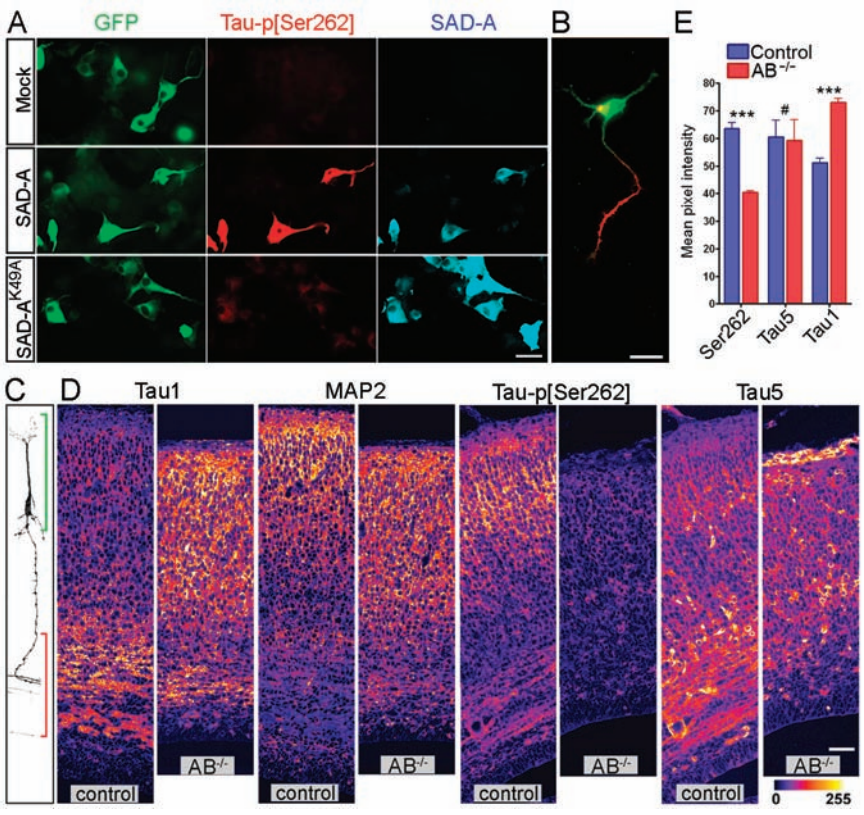


Fig. 4. SAD kinase regulates tau phosphorylation. (**A**) PC12 cells were transfected with a tau-GFP fusion alone (mock) or tau-GFP along with SAD-A or a kinase-defective SAD-A mutant (K49A). Three days later, cells were stained with antibody to SAD-A and antibody to tau-p(S262). SAD-A promotes tau phosphorylation at S262. Scale bar, 100 μm. (**B**) A cultured hippocampal neuron stained with antibody to tau-p(S262) (green) and antibody to dephospho-tau (Tau-1, red), showing concentration of the phosphoepitope in dendrites. Scale bar, 20 μm. (**C**) A single, DiI-labeled pyramidal neuron from a control to indicate the arrangement of neurons; the apical dendrite and axon are bracketed in green and red, respectively. (**D**) Sections of control and $AB^{-/-}$ cortex were labeled with antibodies to dephospho-tau (Tau-1), MAP2, tau-p(S262), or all tau isoforms (Tau5). Intensity is rendered in the color scale at the bottom right. In controls, MAP2 and tau-p(S262) were concentrated in the dendrite-rich upper cortical plate and Tau-1 in the axon-rich intermediate zone. In mutants, levels of dephospho-tau increased and tau-p(S262) decreased in the cortical plate; total tau levels did not change. Scale bar, 50 μm. (**E**) Quantitation of tau levels in the cortical plate. ***, $P < 0.001$; #, $P > 0.8$ by *t* test.

enable them to be locally regulated in neuron-specific ways. It remains to be determined whether mammalian SAD kinases, like *C. elegans* SAD-1 (*5*), are involved not only in polarity but also in synaptogenesis.

References and Notes

1. A. M. Craig, G. Banker, *Annu. Rev. Neurosci.* **17**, 267 (1994).
2. F. Bradke, C. G. Dotti, *Science* **283**, 1931 (1999).
3. M. A. Silverman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7051 (2001).
4. N. Arimura, C. Menager, Y. Fukata, K. Kaibuchi, *J. Neurobiol.* **58**, 34 (2004).
5. J. G. Crump, M. Zhen, Y. Jin, C. I. Bargmann, *Neuron* **29**, 115 (2001).
6. S. Guo, K. J. Kemphues, *Cell* **81**, 611 (1995).
7. A. Wodarz, *Nature Cell Biol.* **4**, E39 (2002).
8. J. R. Sanes, J. W. Lichtman, *Annu. Rev. Neurosci.* **22**, 389 (1999).
9. Y. Sasakura, M. Ogasawara, K. W. Makabe, *Mech. Dev.* **76**, 161 (1998).
10. Materials and methods are available as supporting material on *Science* Online.
11. R. F. Hevner *et al.*, *Dev. Neurosci.* **25**, 139 (2003).
12. O. Marin, J. L. Rubenstein, *Annu. Rev. Neurosci.* **26**, 441 (2003).
13. Y. A. Pan, M. Kishi, J. R. Sanes, in preparation.
14. W. B. Gan, J. Grutzendler, W. T. Wong, R. O. Wong, J. W. Lichtman, *Neuron* **27**, 219 (2000).
15. Z. Molnar, R. Adams, C. Blakemore, *J. Neurosci.* **18**, 5723 (1998).
16. B. Pettmann, C. E. Henderson, *Neuron* **20**, 633 (1998).
17. K. Goslin, H. Asmusse, G. Banker, in *Culturing Nerve Cells*, G. Banker, K. Goslin, Eds. (MIT Press, Cambridge, MA, ed. 2, 1998), pp. 339–370.
18. G. J. Brewer, J. R. Torricelli, E. K. Evege, P. J. Price, *J. Neurosci. Res.* **35**, 567 (1993).
19. G. Drewes, A. Ebneth, U. Preuss, E. M. Mandelkow, E. Mandelkow, *Cell* **89**, 297 (1997).
20. J. W. Mandell, G. A. Banker, *J. Neurosci.* **16**, 5727 (1996).
21. J. Biernat *et al.*, *Mol. Biol. Cell* **13**, 4013 (2002).
22. I. Nishimura, Y. Yang, B. Lu, *Cell* **116**, 671 (2004).
23. J. Biernat, N. Gustke, G. Drewes, E. M. Mandelkow, E. Mandelkow, *Neuron* **11**, 153 (1993).
24. T. Timm *et al.*, *EMBO J.* **22**, 5090 (2003).
25. J. Spicer, A. Ashworth, *Curr. Biol.* **14**, R383 (2004).
26. S. H. Shi, L. Y. Jan, Y. N. Jan, *Cell* **112**, 63 (2003).

27. We thank C. Bargmann for sharing data and helpful comments on the manuscript; R. Burgess, J. Cunningham, I.-J. Kim, T. Kummer, and R. Lewis for advice and assistance; G. Feng for a cDNA library; R. Hevner for antibodies; K. Kosik for green fluorescent protein (GFP)–tau cDNA; and the Mouse Genetics Core at Washington University for blastocyst injections. Supported by grants from NIH (J.R.S.) and an NIH National Research Service Award predoctoral fellowship (Y.A.P).

Supporting Online Material

www.sciencemag.org/cgi/content/full/307/5711/929/DC1
Materials and Methods
Figs. S1 to S5
References and Notes

11 November 2004; accepted 22 December 2004
10.1126/science.1107403

Methylation as a Crucial Step in Plant microRNA Biogenesis

Bin Yu,* Zhiyong Yang,* Junjie Li,† Svetlana Minakhina, Maocheng Yang, Richard W. Padgett, Ruth Steward, Xuemei Chen‡

Methylation on the base or the ribose is prevalent in eukaryotic ribosomal RNAs (rRNAs) and is thought to be crucial for ribosome biogenesis and function. Artificially introduced 2′-*O*-methyl groups in small interfering RNAs (siRNAs) can stabilize siRNAs in serum without affecting their activities in RNA interference in mammalian cells. Here, we show that plant microRNAs (miRNAs) have a naturally occurring methyl group on the ribose of the last nucleotide. Whereas methylation of rRNAs depends on guide RNAs, the methyltransferase protein HEN1 is sufficient to methylate miRNA/miRNA* duplexes. Our studies uncover a new and crucial step in plant miRNA biogenesis and have profound implications in the function of miRNAs.

MicroRNAs are 20- to 24-nucleotide (nt) RNAs that serve as sequence-specific regulators of gene expression in diverse eukaryotic organisms. miRNA/miRNA* duplexes, products of Dicer processing of pre-miRNAs, are structurally similar to small interfering RNAs

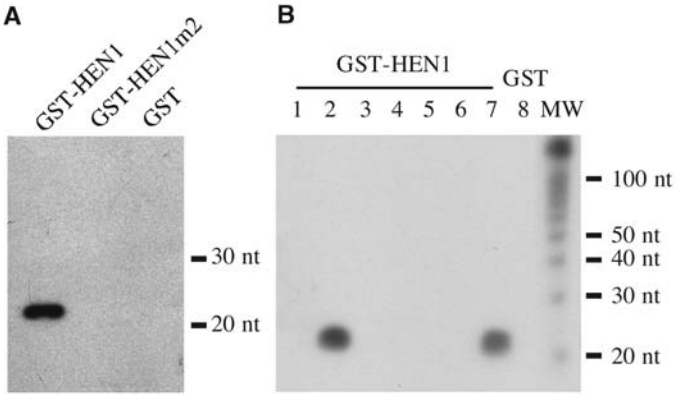


Fig. 1. HEN1 has miRNA methyltransferase activity in vitro. The RNA substrates were incubated with the proteins and resolved by electrophoresis, and the presence of ^{14}C-labeled RNA was detected by autoradiography. (**A**) miRNA methyltransferase assay using the miR173/miR173* duplex as the substrate. (**B**) Various RNA substrates were tested for methylation by GST-HEN1. The substrates used are the putative pre-miR173 (lane 1), miR173/miR173* duplex (lane 2), miR173 (lane 3), miR173* (lane 4), a DNA duplex identical in sequence to miR173/miR173* (lane 5), *E. coli* tRNA (lane 6), and miR167a/miR167a* duplex (lane 7). Lane 8, miR173/miR173* was incubated with GST.

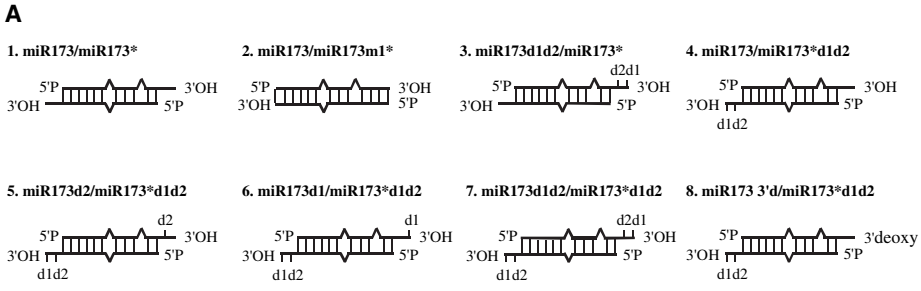
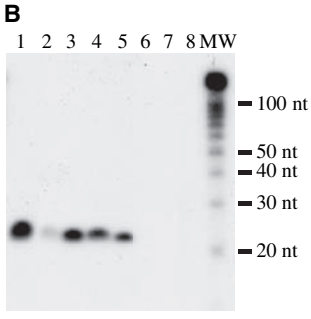


Fig. 2. Methylation of miRNA/miRNA* in vitro requires the two free OH groups on the 3′-most nucleotides. The substrates diagrammed in (**A**) were assayed for methylation by GST-HEN1 in (**B**). The numbers of the molecules in (A) correspond to the lane numbers in (B). In the diagrams, the top strand is miR173 (22 nt), which is 1 nt longer than miR173*, the bottom strand. The two mismatches in the duplex are represented by the bulges. Species 2 in (A) differs from species 1, in that the miR173m1* strand is longer by 2 nt at its 5′ end but shorter by 2 nt at its 3′ end than miR173*. In species 3 to 8, d1 and d2 represent a deoxyribose at the 2′ carbon of the last and penultimate nucleotide, respectively. Species 8 has a deoxyribose at the 3′ carbon of the last nucleotide.



(siRNAs) in that they have 2-nt overhangs, as well as a 5′ phosphate (P) and a 3′ OH (*1*). miRNA accumulation in *Arabidopsis* requires a protein named HEN1 (*2*). In *hen1* mutants, miRNAs are reduced in abundance, and when detectable, are heterogeneous in size (*2*–*5*). The HEN1 protein has an N-terminal putative double-stranded RNA–binding motif and a C-terminal region that is conserved among many bacterial, fungal, and metazoan proteins [(*2*), fig. S1a]. Embedded within this region is a recognizable *S*-adenosyl methionine (SAM)–binding motif (fig. S1a), which suggests that HEN1 may be a miRNA methyltransferase (*6*, *7*).

To test whether HEN1 is a miRNA methyltransferase, we expressed the full-length HEN1 protein fused to glutathione *S*-transferase (GST) in *Escherichia coli* and performed a methylation reaction. In vitro synthesized miR173 and miR173* RNA oligonucleotides with 5′ P and 3′ OH were annealed to generate the miR173/miR173* duplex. Purified GST or GST-HEN1 (fig. S1b) was incubated with the miR173/miR173* duplex in the presence of ^{14}C-labeled SAM. GST-HEN1 indeed resulted in the methylation of the miRNA/miRNA* duplex, but GST alone did not (Fig. 1A). GST-HEN1m2, in which the potential SAM-binding motif in the conserved C-terminal domain of HEN1 was mutated (fig. S1), was unable to methylate miR173/miR173* (Fig. 1A).

We next determined the substrate specificity of HEN1 by testing various molecules as potential substrates in the in vitro reaction. An in vitro transcribed RNA corresponding to the putative pre-miR173 or *E. coli* tRNA was not methylated by GST-HEN1 (Fig. 1B, lanes 1 and 6). Unlike miR173/miR173*, single-stranded miR173 or miR173*, or a DNA duplex with sequences identical to miR173/miR173*, was not methylated by GST-HEN1 (Fig. 1B, lanes 2 to 5). GST-HEN1 could also methylate miR167a/miR167a* (Fig. 1B, lane 7), another miRNA/miRNA* duplex, which suggests that HEN1 does not have sequence specific recognition properties. In addition, two siRNA duplexes with perfect sequence complementarity failed to be methylated by GST-HEN1 (*8*). These results indicate that HEN1 is a methyltransferase and that HEN1 recognizes and acts on miRNA/miRNA* duplexes in vitro. However, although HEN1 is insufficient to methylate siRNA duplexes in vitro, other factors may facilitate the methylation of siRNAs in vivo. In fact, *HEN1* is required for the accumulation of siRNAs from sense transgenes (*5*).

Where does HEN1 methylate miRNA/miRNA* in vitro? miR173/miR173m1* (species 2 in Fig. 2A), a duplex without any 3′ overhangs, was a poor substrate for HEN1 (Fig. 2B, compare lane 2 with lane 1), which suggests that the 2 nt 3′ overhangs are an important feature in the substrate. In fact, HEN1 requires the 2′ OH on at least one of the last two nucleotides in the substrate because miR173d1d2/miR173*d1d2, a duplex identical to miR173/miR173* in sequence but with 2′-deoxyribose in the last two nucleotides (Fig. 2A, species 7), failed to be methylated (Fig. 2B, lane 7). Both miR173d1d2/miR173* and miR173/miR173*d1d2 (species 3 and 4 in Fig. 2A) could be methylated by HEN1 (Fig. 2B, lanes 3 and 4), which suggests that HEN1, at least in vitro, does not distinguish the miRNA strand from miRNA*. miR173d2/miR173*d1d2 (Fig. 2A, species 5), in which the penultimate nucleotide of miR173 is a deoxynucleotide, was still methylated by HEN1 (Fig. 2B, lane 5). However, miR173d1/miR173*d1d2 (Fig. 2A, species 6), in which the last nucleotide of miR173 is a deoxynucleotide at the 2′ position of the ribose, failed to be methylated by HEN1 (Fig. 2B, lane 6). Therefore, the 2′ OH on the last nucleotide is critical for the methylation reaction. The last nucleotide differs from all other nucleotides in that it also has a 3′ OH. A duplex identical to miR173/miR173*d1d2 except that the last nucleotide of miR173 is a deoxynucleotide at the 3′ position of the ribose (Fig. 2A, species 8) also failed to be methylated (Fig. 2B, lane 8), which indicated that the 3′ OH is also critical for the methylation reaction.

The OH groups on the last nucleotide may be the actual sites of methylation or

Waksman Institute, Rutgers University, Piscataway, NJ 08854, USA.

*These authors contributed equally to this work.
†Present address: Howard Hughes Medical Institute and Skirball Institute of Biomolecular Medicine, New York University School of Medicine, 540 First Avenue, Skirball 2-17, New York, NY 10016, USA.
‡To whom correspondence should be addressed. E-mail: xuemei@waksman.rutgers.edu

may only be features necessary for substrate recognition. We tested whether methylation occurs on at least one of the OH on the last nucleotide. After incubation with GST-HEN1 in the presence of [^{14}C]SAM, the miR173/miR173* duplex was treated with sodium periodate followed by β elimination, chemical reactions that require the presence of OH groups at both 2′ and 3′ positions on the ribose of the last nucleotide. If the chemical reactions occur, the miRNA will be one nucleotide shorter and have a 3′ P (*9*), so that the apparent mobility would be increased by two nucleotides (*10*). The effectiveness of the chemical reactions was demonstrated with control miR173 oligonucleotides with or without a 2′-*O*-methyl on its last nucleotide (fig. S2a). A mobility shift after the chemical reactions was observed for miR173 incubated with GST alone, but not for the ^{14}C-labeled miRNA resulting from the incubation with GST-HEN1 (fig. S2b), which indicated that methylation by GST-HEN1 had occurred on at least one of the two OH groups on the last nucleotide of miR173.

Are plant miRNAs methylated on the ribose of the last nucleotide as predicted from our in vitro results? We first performed the oxidation and β elimination reactions to evaluate whether the 2′ or 3′ OH on the 3′-most nucleotides is blocked in plant miRNAs. Total RNAs were treated with the chemicals, and specific miRNAs were detected by filter hybridization. In vitro synthesized miR173-OH (without any methyl groups) or miR173-2′me (with a 2′-*O*-methyl on its last nucleotide) was mixed with total RNAs from *dcl1-9*, which does not have detectable levels of miR173 (*2*), to show that the treatments were effective (Fig. 3A). The endogenous miR173 in wild type did not show any mobility shift (Fig. 3B). The miR173-OH standard was also included in the total RNA from wild type to demonstrate that the treatment was sufficient to induce the mobility shift (Fig. 3B). Similar results were obtained for miR167, miR172, miR159, and miR163 (Fig. 3C). Therefore, the last nucleotide of *Arabidopsis* miRNAs is blocked at the 2′-*O*-ribose, 3′-*O*-ribose, or both positions. Furthermore, for each miRNA species, any unmethylated miRNAs, if present, were below the limits of detection (fig. S3). In the *hen1-1* mutant, the heterogeneous miRNAs, when detectable, showed a mobility shift after the oxidation and β elimination reactions (Fig. 3C), which indicates that HEN1 is responsible for blocking the OH group(s) in vivo.

To determine the nature and number of modifications that are present in plant miRNAs, we used affinity purification to isolate miR173 from *Arabidopsis* total RNAs with a biotinylated antisense DNA oligonucleotide in order to perform mass spectrometry to determine the molecular mass of this miRNA. miR173 was chosen because no related miRNA genes are present in the genome. The final RNA fraction from the purification contained miR173 and was free of contamination from other miRNA species (fig. S4, a and c). The fraction, however, did contain trace amounts of larger RNAs that comigrated with tRNAs or rRNAs (fig. S4d) and some antisense DNA that was used as the affinity probe (fig. S4b). The amount of the DNA was expected to be below the limits of detection by our mass spectrometry analysis. In addition, we designed the DNA oligonucleotide with two extra nucleotides, so that the expected molecular mass would differ greatly from that of miR173. In the mass spectrometry analysis of the final purified RNA fraction, a large peak of unresolved molecular mass was detected and probably corresponds to the contaminating large RNAs in the fraction (fig. S5a). In the resolvable mass range (less than 8000 dalton), only one major peak with a mass of 7118.9 dalton was found (Fig. 4A; fig. S5a). This peak could not be the DNA oligonucleotide because the expected molecular mass of the DNA is 7758.3 dalton. Therefore, this peak must correspond to the in vivo miR173 species. The mass of the in vivo miR173 is larger than the expected mass of miR173 standard without any modifications (fig. S5b) by 13.6 dalton and is close to that of the miR173 standard with one methyl group (Fig. 4B). The deviation of the measured mass of the in vivo miR173 from the expected mass of the 1 methyl standard fell well within the experimental error (0.025%), which was measured with miR173 standards harboring 0, 1, or 2 methyl groups (Fig. 4B; fig. S5, b and c). Therefore, we conclude that miR173 from *Arabidopsis* has one methyl group, which appears to be the only form of modification on the bulk of miR173. This and results of the β elimination reactions indicated that the methyl group is either at the 2′ or the 3′-*O*-ribose position of the last nucleotide.

Although proteins homologous to HEN1 along its entire length are not found in non-

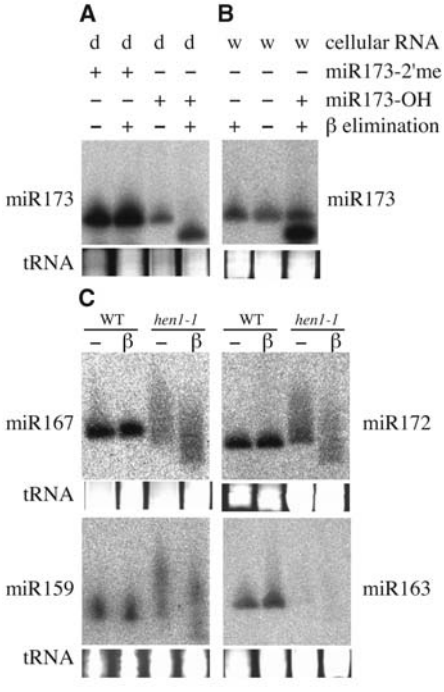


Fig. 3. One or both OH groups on the ribose of the last nucleotide of *Arabidopsis* miRNAs are blocked. Total RNAs were treated with periodate followed by β elimination and then subjected to RNA filter hybridization with probes against various miRNAs (as indicated next to the blots). The region of the stained gel corresponding to where tRNAs migrate is shown below the hybridization images to indicate the amount of total RNAs used. (**A**) In vitro synthesized miR173-OH or miR173-2′me was added to total RNA from *dcl1-9* (d) and treated with the chemicals. miR173-OH showed the expected mobility shift but miR173-2′me was resistant to the chemical modification reactions. (**B**) miR173 from wild-type total RNA (w) is resistant to the chemical modification reactions. (**C**) Total RNA from wild type (wt) or *hen1-1* was either subjected (β) or not (–) to the chemical modification reactions and probed for various miRNAs.

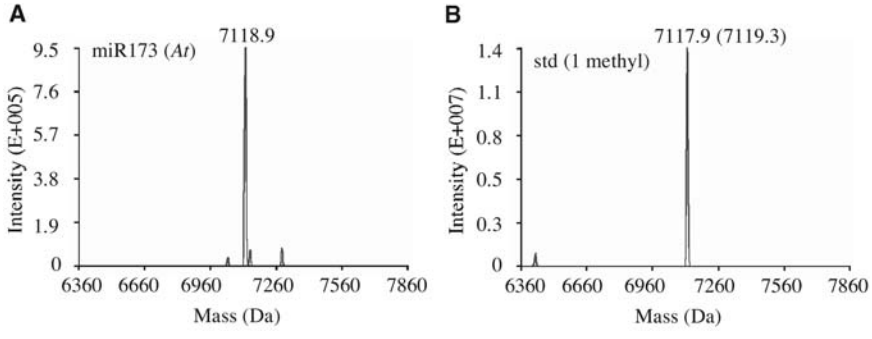


Fig. 4. Mass spectrometry analysis of miR173 isolated from *Arabidopsis* (**A**) and a miR173 standard with one methyl group (**B**). The position of a peak along the *x* axis represents the molecular mass of the species. Signal intensity approximately reflects the amount of the RNA in each peak. The expected mass of the standard is indicated by the number in parentheses. The measured mass is indicated by the numbers directly above the peaks.

plant organisms, ones homologous to the C-terminal methyltransferase domain of HEN1 are present in diverse organisms (*2*). We examined three miRNAs from *C. elegans* and four from *Drosophila* and found that they have free 2′-OH and 3′-OH groups on their last nucleotides (fig. S6). *Drosophila let-7* was previously shown to have free 2′ and 3′ OH on its last nucleotide (*9*). Therefore, animal miRNAs may not be modified as they are in plants. However, it cannot be ruled out that methylation of animal miRNAs occurs in a small temporal or spatial window during development.

We envision several potential functions of the methyl group in plant miRNAs. First, it may protect miRNAs by attenuating enzymes such as exonucleases that target the 3′ end of miRNAs. This would be consistent with the reduced abundance of miRNAs in *hen1* mutants. Second, the methyl group may facilitate the recognition of miRNAs by plant argonaute proteins in RNA-induced silencing complex (RISC) assembly. In fact, the 2′ OH of the last nucleotide is a site of contact by the PAZ domain in animal argonaute proteins (*11*, *12*), and 2′-*O*-methyl on the last nucleotide decreases the binding affinity by the PAZ domain (*12*). Perhaps plant argonaute proteins, unlike their animal counterparts, prefer methylated miRNAs. It is noteworthy that previous efforts to program RISC with exogenously introduced miRNAs in plant extracts have been unsuccessful (*13*). We suspect this lack of success was due to the inefficient incorporation of unmethylated miRNAs into RISC. Third, the methylation may limit the ability of plant miRNAs to serve as primers for RNA-dependent RNA polymerase (RdRP). Plant miRNAs show extensive complementarity to their target mRNAs, which raises the possibility that miRNAs may serve as primers to allow RdRP to generate dsRNA using the target mRNA as a template. The presence of the methyl group may reduce the efficiency of such a reaction. Note that miRNA-mediated cleavage of target mRNAs can lead to transitive RNA silencing dependent on an RdRP (*14*) and that HEN1 is required for sense transgene silencing that also depends on an RdRP (*15*, *16*). However, the generation of dsRNAs by RdRP in these processes does not necessarily have to be primed by small RNAs (*13*).

References and Notes

1. D. P. Bartel, *Cell* **116**, 281 (2004).
2. W. Park, J. Li, R. Song, J. Messing, X. Chen, *Curr. Biol.* **12**, 1484 (2002).
3. Z. Xie *et al.*, *PLoS Biol.* **2**, E104 (2004).
4. M. H. Han, S. Goud, L. Song, N. Fedoroff, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 1093 (2004).
5. S. Boutet *et al.*, *Curr. Biol.* **13**, 843 (2003).
6. V. Anantharaman, E. V. Koonin, L. Aravind, *Nucleic Acids Res.* **30**, 1427 (2002).
7. A. C. Mallory, H. Vaucheret, *Curr. Opin. Plant Biol.* **7**, 120 (2004).
8. Z. Yang, X. Chen, unpublished observations.
9. S. Alefelder, B. K. Patel, F. Eckstein, *Nucleic Acids Res.* **26**, 4983 (1998).
10. G. Hutvágner *et al.*, *Science* **293**, 834 (2001).
11. A. Lingel, B. Simon, E. Izaurralde, M. Sattler, *Nat. Struct. Mol. Biol.* **11**, 576 (2004).
12. J. B. Ma, K. Ye, D. J. Patel, *Nature* **429**, 318 (2004).
13. G. Tang, B. J. Reinhart, D. P. Bartel, P. D. Zamore, *Genes Dev.* **17**, 49 (2003).
14. E. A. Parizotto, P. Dunoyer, N. Rahm, C. Himber, O. Voinnet, *Genes Dev.* **18**, 2237 (2004).
15. P. Mourrain *et al.*, *Cell* **101**, 533 (2000).
16. T. Dalmay, A. Hamilton, S. Rudd, S. Angell, D. C. Baulcombe, *Cell* **101**, 543 (2000).

17. We are grateful to J. Liu for antibodies against GST-HEN1 and for discussions on the project. We thank Y. Ebright for suggestions on nucleotide chemistry and L. Zhao, Y.-J. Kim, and S. Mlotshwa for comments on the manuscript. This work was supported by NIH (GM61146) and NSF (MCB-0343480) grants to X.C., a grant from the NIH and an academic excellence grant from Rutgers University to R.W.P. and a grant from NIH to R.S.

Supporting Online Material

www.sciencemag.org/cgi/content/full/307/5711/932/DC1
Materials and Methods
Figs. S1 to S6
References and Notes

4 November 2004; accepted 13 December 2004
10.1126/science.1107130

A Selective Inhibitor of eIF2α Dephosphorylation Protects Cells from ER Stress

Michael Boyce,[1] Kevin F. Bryant,[2]* Céline Jousse,[3]* Kai Long,[4]* Heather P. Harding,[3] Donalyn Scheuner,[5] Randal J. Kaufman,[5] Dawei Ma,[4] Donald M. Coen,[2] David Ron,[3] Junying Yuan[1]†

Most protein phosphatases have little intrinsic substrate specificity, making selective pharmacological inhibition of specific dephosphorylation reactions a challenging problem. In a screen for small molecules that protect cells from endoplasmic reticulum (ER) stress, we identified salubrinal, a selective inhibitor of cellular complexes that dephosphorylate eukaryotic translation initiation factor 2 subunit α (eIF2α). Salubrinal also blocks eIF2α dephosphorylation mediated by a herpes simplex virus protein and inhibits viral replication. These results suggest that selective chemical inhibitors of eIF2α dephosphorylation may be useful in diseases involving ER stress or viral infection. More broadly, salubrinal demonstrates the feasibility of selective pharmacological targeting of cellular dephosphorylation events.

All eukaryotes respond to ER stress through a set of pathways known as the unfolded protein response (UPR) (*1*, *2*). The UPR is essential for cellular homeostasis, and its dysregulation has been implicated in many important pathologies, including diabetes, Alzheimer's disease, and viral infection (*3*). However, several aspects of the mammalian UPR remain obscure. In particular, little is known about the control of the mammalian apoptotic pathways activated by excessive, uncorrected ER stress (*4*) or how these pathways might be manipulated for therapeutic benefit. We used a chemical biology approach to study apoptosis induced by ER stress in mammalian cells.

In screening ~19,000 chemicals for compounds that protect the rat pheochromocytoma cell line PC12 from ER stress-induced apoptosis, we identified a small molecule we termed salubrinal (sal) (Fig. 1A). Sal inhibited ER stress-mediated apoptosis induced by the protein glycosylation inhibitor tunicamycin (Tm) in a dose-dependent manner, with a median effective concentration (EC_{50}) ~ 15 μM (Fig. 1B). Cytoprotection by sal exceeded that by zVAD.fmk, a pan-inhibitor of the caspase family of apoptotic proteases (Fig. 1B). Sal also suppressed Tm-induced DNA fragmentation (Fig. 1C), the processing of caspase-7, a caspase activated by ER stress (Fig. 1D) (*5*), and the activity of caspase-3- and caspase-7-type enzymes (fig. S1A), confirming its anti-apoptotic effect. Cytoprotection was not specific to Tm, because sal protected cells from brefeldin A, which causes ER stress by blocking ER-to-Golgi vesicle transport (fig. S1B). However, sal is not a general apoptosis inhibitor, because it did not protect against apoptotic stimuli unrelated to ER stress (fig. S1, C and D). Importantly, sal was nontoxic at concentrations required for maximal cytoprotection (TD_{50} >

[1]Department of Cell Biology and [2]Department of Biological Chemistry and Molecular Pharmacology and Committee on Virology, Harvard Medical School, Boston, MA 02115, USA. [3]Skirball Institute, New York University School of Medicine, New York, NY 10016, USA. [4]Shanghai Institute of Organic Chemistry, Shanghai, China. [5]Department of Biological Chemistry and Howard Hughes Medical Institute, University of Michigan Medical Center, Ann Arbor, MI 48109, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: jyuan@hms.harvard.edu

plant organisms, ones homologous to the C-terminal methyltransferase domain of HEN1 are present in diverse organisms (*2*). We examined three miRNAs from *C. elegans* and four from *Drosophila* and found that they have free 2′-OH and 3′-OH groups on their last nucleotides (fig. S6). *Drosophila let-7* was previously shown to have free 2′ and 3′ OH on its last nucleotide (*9*). Therefore, animal miRNAs may not be modified as they are in plants. However, it cannot be ruled out that methylation of animal miRNAs occurs in a small temporal or spatial window during development.

We envision several potential functions of the methyl group in plant miRNAs. First, it may protect miRNAs by attenuating enzymes such as exonucleases that target the 3′ end of miRNAs. This would be consistent with the reduced abundance of miRNAs in *hen1* mutants. Second, the methyl group may facilitate the recognition of miRNAs by plant argonaute proteins in RNA-induced silencing complex (RISC) assembly. In fact, the 2′ OH of the last nucleotide is a site of contact by the PAZ domain in animal argonaute proteins (*11*, *12*), and 2′-*O*-methyl on the last nucleotide decreases the binding affinity by the PAZ domain (*12*). Perhaps plant argonaute proteins, unlike their animal counterparts, prefer methylated miRNAs. It is noteworthy that previous efforts to program RISC with exogenously introduced miRNAs in plant extracts have been unsuccessful (*13*). We suspect this lack of success was due to the inefficient incorporation of unmethylated miRNAs into RISC. Third, the methylation may limit the ability of plant miRNAs to serve as primers for RNA-dependent RNA polymerase (RdRP). Plant miRNAs show extensive complementarity to their target mRNAs, which raises the possibility that miRNAs may serve as primers to allow RdRP to generate dsRNA using the target mRNA as a template. The presence of the methyl group may reduce the efficiency of such a reaction. Note that miRNA-mediated cleavage of target mRNAs can lead to transitive RNA silencing dependent on an RdRP (*14*) and that HEN1 is required for sense transgene silencing that also depends on an RdRP (*15*, *16*). However, the generation of dsRNAs by RdRP in these processes does not necessarily have to be primed by small RNAs (*13*).

References and Notes

1. D. P. Bartel, *Cell* **116**, 281 (2004).
2. W. Park, J. Li, R. Song, J. Messing, X. Chen, *Curr. Biol.* **12**, 1484 (2002).
3. Z. Xie *et al.*, *PLoS Biol.* **2**, E104 (2004).
4. M. H. Han, S. Goud, L. Song, N. Fedoroff, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 1093 (2004).
5. S. Boutet *et al.*, *Curr. Biol.* **13**, 843 (2003).
6. V. Anantharaman, E. V. Koonin, L. Aravind, *Nucleic Acids Res.* **30**, 1427 (2002).
7. A. C. Mallory, H. Vaucheret, *Curr. Opin. Plant Biol.* **7**, 120 (2004).
8. Z. Yang, X. Chen, unpublished observations.
9. S. Alefelder, B. K. Patel, F. Eckstein, *Nucleic Acids Res.* **26**, 4983 (1998).
10. G. Hutvágner *et al.*, *Science* **293**, 834 (2001).
11. A. Lingel, B. Simon, E. Izaurralde, M. Sattler, *Nat. Struct. Mol. Biol.* **11**, 576 (2004).
12. J. B. Ma, K. Ye, D. J. Patel, *Nature* **429**, 318 (2004).
13. G. Tang, B. J. Reinhart, D. P. Bartel, P. D. Zamore, *Genes Dev.* **17**, 49 (2003).
14. E. A. Parizotto, P. Dunoyer, N. Rahm, C. Himber, O. Voinnet, *Genes Dev.* **18**, 2237 (2004).
15. P. Mourrain *et al.*, *Cell* **101**, 533 (2000).
16. T. Dalmay, A. Hamilton, S. Rudd, S. Angell, D. C. Baulcombe, *Cell* **101**, 543 (2000).

17. We are grateful to J. Liu for antibodies against GST-HEN1 and for discussions on the project. We thank Y. Ebright for suggestions on nucleotide chemistry and L. Zhao, Y.-J. Kim, and S. Mlotshwa for comments on the manuscript. This work was supported by NIH (GM61146) and NSF (MCB-0343480) grants to X.C., a grant from the NIH and an academic excellence grant from Rutgers University to R.W.P. and a grant from NIH to R.S.

Supporting Online Material

www.sciencemag.org/cgi/content/full/307/5711/932/DC1
Materials and Methods
Figs. S1 to S6
References and Notes

4 November 2004; accepted 13 December 2004
10.1126/science.1107130

A Selective Inhibitor of eIF2α Dephosphorylation Protects Cells from ER Stress

Michael Boyce,[1] Kevin F. Bryant,[2]* Céline Jousse,[3]* Kai Long,[4]* Heather P. Harding,[3] Donalyn Scheuner,[5] Randal J. Kaufman,[5] Dawei Ma,[4] Donald M. Coen,[2] David Ron,[3] Junying Yuan[1]†

Most protein phosphatases have little intrinsic substrate specificity, making selective pharmacological inhibition of specific dephosphorylation reactions a challenging problem. In a screen for small molecules that protect cells from endoplasmic reticulum (ER) stress, we identified salubrinal, a selective inhibitor of cellular complexes that dephosphorylate eukaryotic translation initiation factor 2 subunit α (eIF2α). Salubrinal also blocks eIF2α dephosphorylation mediated by a herpes simplex virus protein and inhibits viral replication. These results suggest that selective chemical inhibitors of eIF2α dephosphorylation may be useful in diseases involving ER stress or viral infection. More broadly, salubrinal demonstrates the feasibility of selective pharmacological targeting of cellular dephosphorylation events.

All eukaryotes respond to ER stress through a set of pathways known as the unfolded protein response (UPR) (*1*, *2*). The UPR is essential for cellular homeostasis, and its dysregulation has been implicated in many important pathologies, including diabetes, Alzheimer's disease, and viral infection (*3*). However, several aspects of the mammalian UPR remain obscure. In particular, little is known about the control of the mammalian apoptotic pathways activated by excessive, uncorrected ER stress (*4*) or how these pathways might be manipulated for therapeutic benefit. We used a chemical biology approach to study apoptosis induced by ER stress in mammalian cells.

In screening ~19,000 chemicals for compounds that protect the rat pheochromocytoma cell line PC12 from ER stress-induced apoptosis, we identified a small molecule we termed salubrinal (sal) (Fig. 1A). Sal inhibited ER stress-mediated apoptosis induced by the protein glycosylation inhibitor tunicamycin (Tm) in a dose-dependent manner, with a median effective concentration (EC_{50}) ~ 15 μM (Fig. 1B). Cytoprotection by sal exceeded that by zVAD.fmk, a pan-inhibitor of the caspase family of apoptotic proteases (Fig. 1B). Sal also suppressed Tm-induced DNA fragmentation (Fig. 1C), the processing of caspase-7, a caspase activated by ER stress (Fig. 1D) (*5*), and the activity of caspase-3- and caspase-7-type enzymes (fig. S1A), confirming its anti-apoptotic effect. Cytoprotection was not specific to Tm, because sal protected cells from brefeldin A, which causes ER stress by blocking ER-to-Golgi vesicle transport (fig. S1B). However, sal is not a general apoptosis inhibitor, because it did not protect against apoptotic stimuli unrelated to ER stress (fig. S1, C and D). Importantly, sal was nontoxic at concentrations required for maximal cytoprotection (TD_{50} >

[1]Department of Cell Biology and [2]Department of Biological Chemistry and Molecular Pharmacology and Committee on Virology, Harvard Medical School, Boston, MA 02115, USA. [3]Skirball Institute, New York University School of Medicine, New York, NY 10016, USA. [4]Shanghai Institute of Organic Chemistry, Shanghai, China. [5]Department of Biological Chemistry and Howard Hughes Medical Institute, University of Michigan Medical Center, Ann Arbor, MI 48109, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: jyuan@hms.harvard.edu

100 μM) (Fig. 1, B and C), making it a useful tool for studying ER stress.

To determine how sal protects cells from ER stress, we examined its effects on known components of the UPR. Unlike Tm, sal did not up-regulate the canonical UPR targets Xbp-1, Grp78/Bip, and Grp94 (*6*–*8*) (fig. S2A), indicating that sal does not cause ER stress or activate the transcription-dependent branch of the UPR (*2*). A second branch of the UPR involves phosphorylation of eIF2α on Ser^{51} by the ER-localized kinase PERK (*9*, *10*). Phosphorylated eIF2α mediates both a transient decrease in global translation and the translational up-regulation of selected stress-induced mRNAs (*11*). Like Tm, sal induced rapid and robust eIF2α phosphorylation (Fig. 2A and fig. S2, B to D) and its downstream effects, including down-regulation of cyclin D1 (*12*) and up-regulation of GADD34 (*13*) and CHOP (*11*), two proteins whose expression is induced by eIF2α phosphorylation (fig. S2C). Thus, sal selectively engages the translational control branch of the UPR by inducing eIF2α phosphorylation without affecting the transcription-dependent component of the UPR.

EIF2α phosphorylation is cytoprotective during ER stress, because cells are sensitized when this pathway is genetically ablated (*14*, *15*) and protected when it is ectopically enforced (*16*, *17*). We therefore asked whether the induction of eIF2α phosphorylation by sal was related to its anti-apoptotic effects. Sal-induced eIF2α phosphorylation was dose-dependent (Fig. 2A) and correlated well with its protection from Tm (Fig. 1B). Furthermore, chemical derivatives of sal that protected cells from Tm, such as **2**, also induced eIF2α phosphorylation (Fig. 2B), whereas derivatives that failed to protect cells, such as **3**, did not (Fig. 2C). Taken together with previous work establishing the protective function of eIF2α phosphorylation, these results suggest that eIF2α phosphorylation induction by sal accounts at least in part for its anti-apoptotic activity during ER stress.

To determine how sal induces eIF2α phosphorylation, we asked whether sal activates one of the four mammalian eIF2α kinases: PERK, GCN2, RNA-activated protein kinase (PKR), and HRI (*9*, *18*–*20*). Consistent with previous reports (*14*), $PERK^{-/-}$ mouse embryo fibroblasts (MEFs) did not phosphorylate eIF2α

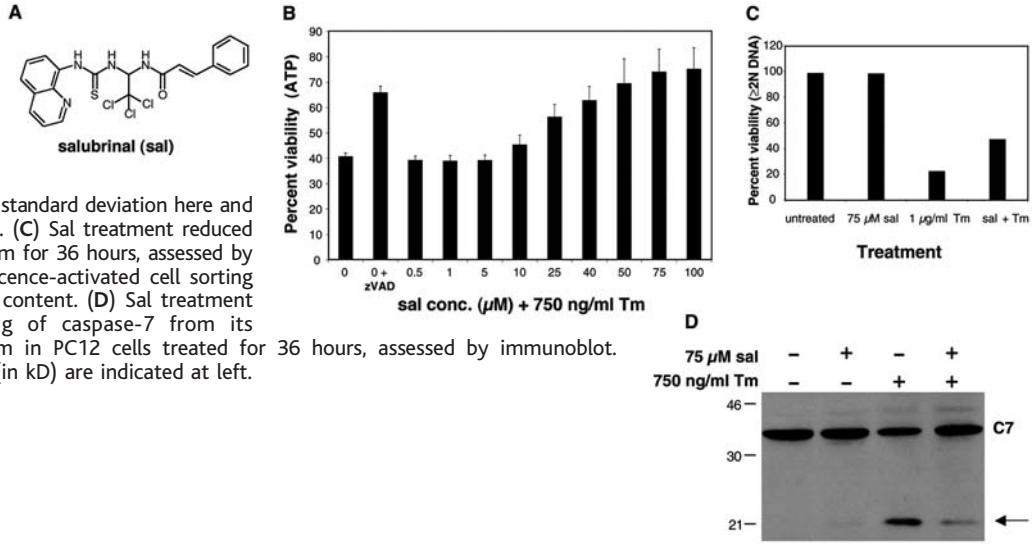


Fig. 1. Sal protects cells against ER stress-induced apoptosis. (**A**) Sal. (**B**) Dose-dependent protection by sal of PC12 cells treated with Tm and concentrations (conc.) of sal as indicated for 37 hours, assessed by cellular adenosine triphosphate (ATP) content. 100 μM zVAD.fmk serves as a positive control. Error bars represent standard deviation here and throughout unless otherwise indicated. (**C**) Sal treatment reduced apoptosis in PC12 cells treated with Tm for 36 hours, assessed by propidium iodide staining and fluorescence-activated cell sorting of cells with subdiploid (< 2N) DNA content. (**D**) Sal treatment reduced the Tm-induced processing of caspase-7 from its zymogen (C7) to active (arrow) form in PC12 cells treated for 36 hours, assessed by immunoblot. Positions of molecular mass markers (in kD) are indicated at left.

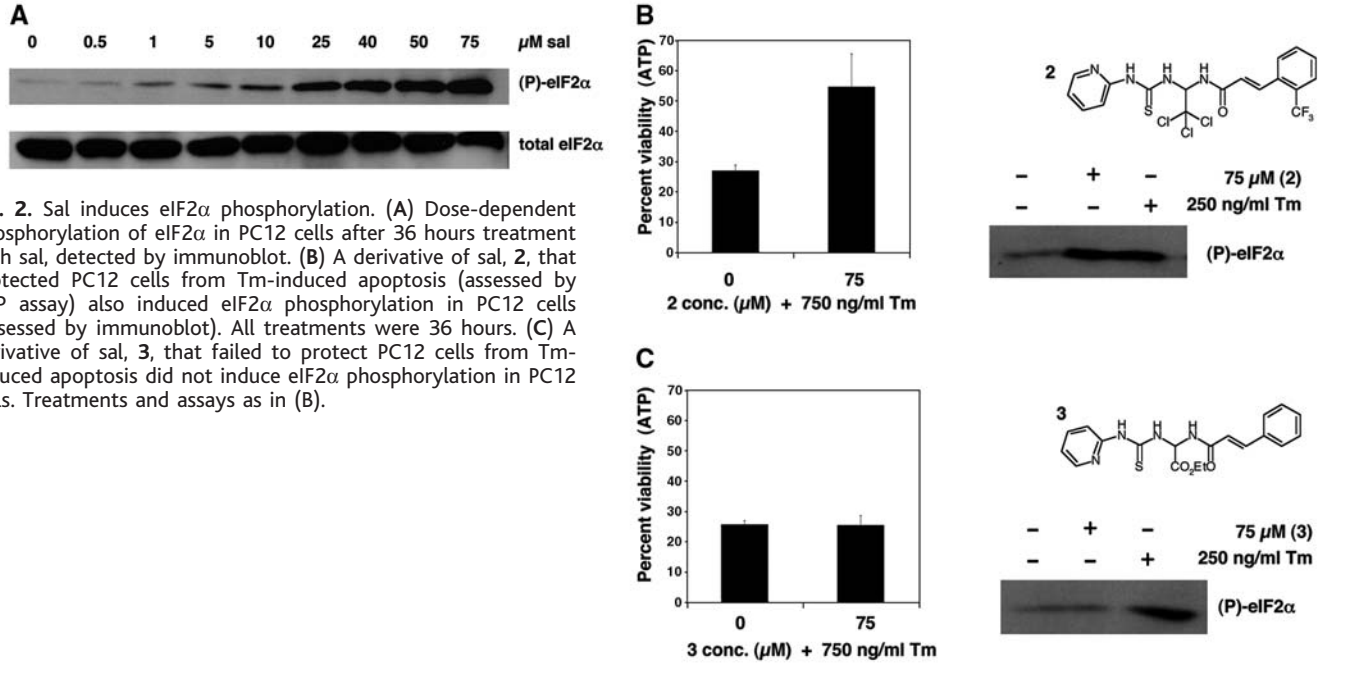


Fig. 2. Sal induces eIF2α phosphorylation. (**A**) Dose-dependent phosphorylation of eIF2α in PC12 cells after 36 hours treatment with sal, detected by immunoblot. (**B**) A derivative of sal, **2**, that protected PC12 cells from Tm-induced apoptosis (assessed by ATP assay) also induced eIF2α phosphorylation in PC12 cells (assessed by immunoblot). All treatments were 36 hours. (**C**) A derivative of sal, **3**, that failed to protect PC12 cells from Tm-induced apoptosis did not induce eIF2α phosphorylation in PC12 cells. Treatments and assays as in (B).

(*21*) or induce the eIF2α phosphorylation–dependent expression of CHOP in response to Tm treatment (fig. S3). However, sal induced eIF2α phosphorylation (*21*) and CHOP expression (fig. S3) regardless of PERK genotype, demonstrating that sal action does not depend on PERK. Similarly, sal induced eIF2α phosphorylation in MEFs deficient in GCN2, PKR, or HRI (*21*), indicating that sal action does not depend on any single eIF2α kinase.

Next, we asked whether sal inhibited eIF2α dephosphorylation. During ER stress, eIF2α is dephosphorylated by a complex containing the serine/threonine phosphatase PP1 and its nonenzymatic cofactor GADD34 (*13*). In an in vitro dephosphorylation assay (*13*), lysate from cells overexpressing an active fragment of GADD34 ($GADD34^C$) efficiently dephosphorylated eIF2α, whereas lysate from vector-only control cells did not (fig. S4A). However, when $GADD34^C$-overexpressing cells were treated with sal, eIF2α dephosphorylation by the resulting lysate was inhibited (Fig. 3A). CReP, a recently discovered homolog of GADD34, is constitutively expressed and mediates eIF2α dephosphorylation in unstressed cells (*16*). Sal also inhibited in vitro dephosphorylation of eIF2α by lysate from CReP-overexpressing cells (fig. S5A) and induced eIF2α phosphorylation in $GADD34^{-/-}$ cells (fig. S5B), where CReP is active (*16*). Therefore, sal inhibits eIF2α dephosphorylation mediated by both constitutive and ER stress-induced phosphatase complexes.

We tested the impact of sal on eIF2α dephosphorylation inside intact cells with the use of a CHOP–green fluorescent protein (GFP) reporter transgene whose expression is stringently dependent upon eIF2α phosphorylation (*22*). Both sal and Tm induced the CHOP-GFP reporter to a similar extent (Fig. 3B). However, stable overexpression of GADD34 nearly abrogated CHOP-GFP induction by Tm (Fig. 3B), because increased GADD34/PP1 phosphatase activity removes the Tm-induced eIF2α phosphorylation. In contrast, sal markedly activated CHOP-GFP even in the presence of overexpressed GADD34 (Fig. 3B) or CReP (fig. S5C). Therefore, sal is mechanistically distinct from Tm because sal can counteract enforced eIF2α phosphatase activ-

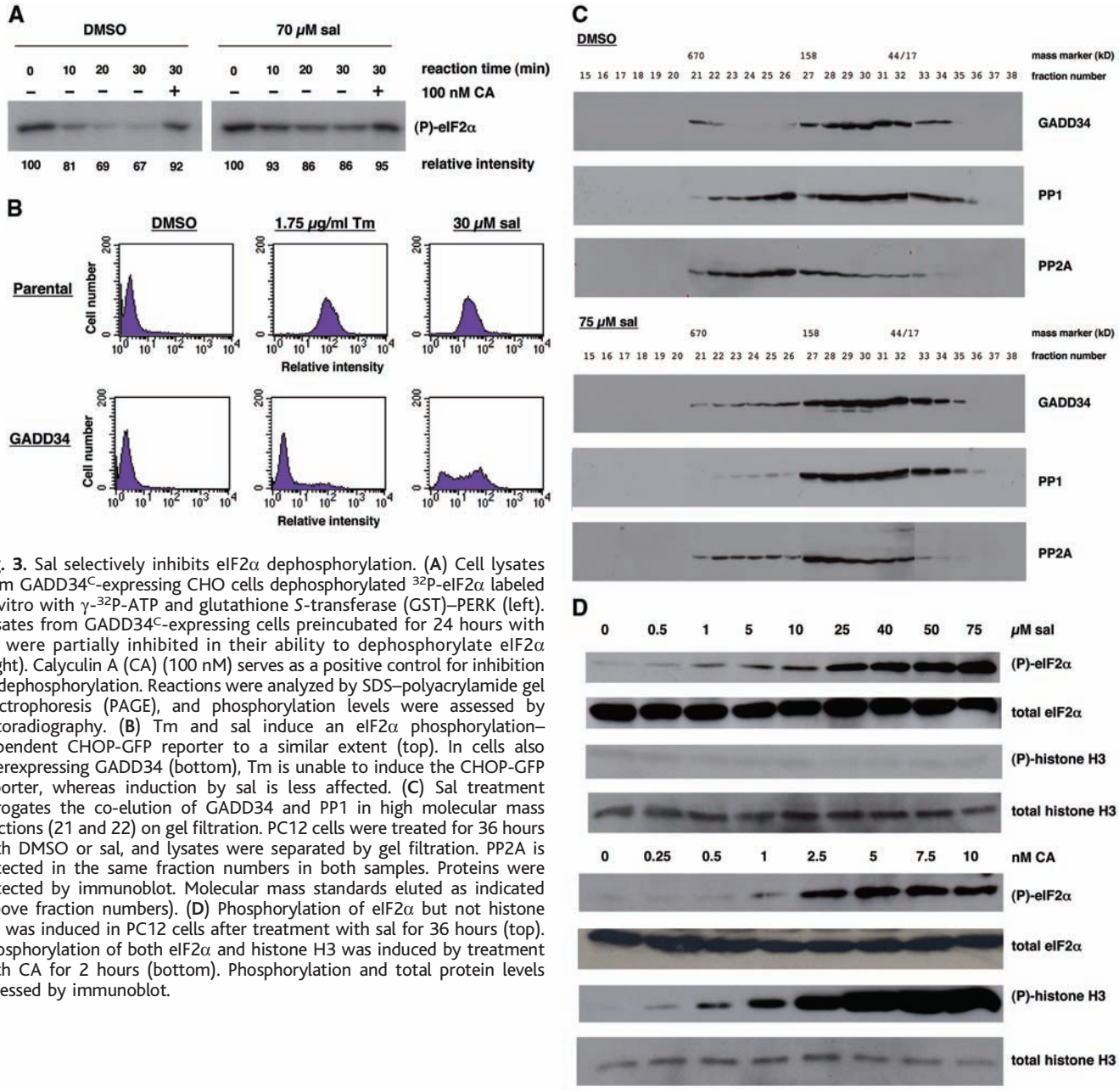


Fig. 3. Sal selectively inhibits eIF2α dephosphorylation. (**A**) Cell lysates from $GADD34^C$-expressing CHO cells dephosphorylated ^{32}P-eIF2α labeled in vitro with γ-^{32}P-ATP and glutathione *S*-transferase (GST)–PERK (left). Lysates from $GADD34^C$-expressing cells preincubated for 24 hours with sal were partially inhibited in their ability to dephosphorylate eIF2α (right). Calyculin A (CA) (100 nM) serves as a positive control for inhibition of dephosphorylation. Reactions were analyzed by SDS–polyacrylamide gel electrophoresis (PAGE), and phosphorylation levels were assessed by autoradiography. (**B**) Tm and sal induce an eIF2α phosphorylation–dependent CHOP-GFP reporter to a similar extent (top). In cells also overexpressing GADD34 (bottom), Tm is unable to induce the CHOP-GFP reporter, whereas induction by sal is less affected. (**C**) Sal treatment abrogates the co-elution of GADD34 and PP1 in high molecular mass fractions (21 and 22) on gel filtration. PC12 cells were treated for 36 hours with DMSO or sal, and lysates were separated by gel filtration. PP2A is detected in the same fraction numbers in both samples. Proteins were detected by immunoblot. Molecular mass standards eluted as indicated (above fraction numbers). (**D**) Phosphorylation of eIF2α but not histone H3 was induced in PC12 cells after treatment with sal for 36 hours (top). Phosphorylation of both eIF2α and histone H3 was induced by treatment with CA for 2 hours (bottom). Phosphorylation and total protein levels assessed by immunoblot.

ity. However, CHOP-GFP induction by sal was antagonized by GADD34 or CReP overexpression, consistent with the model that these are the relevant targets of sal.

As a first step to characterize the sal mechanism, we examined its effects on the GADD34/PP1 complex. GADD34 and PP1 were present in the same high molecular mass gel filtration fractions in lysate from dimethyl sulfoxide (DMSO)–treated cells but not in lysate from sal-treated cells (Fig. 3C, fractions 21 and 22). Thus, sal may affect the composition of an active GADD34/PP1 complex inside cells. As a specificity control, the serine/threonine phosphatase PP2A was detectable in the same fractions regardless of sal treatment (Fig. 3C). Along with the in vitro dephosphorylation and CHOP-GFP experiments, these results indicate that sal inhibits eIF2α dephosphorylation by cellular complexes. Whether sal inhibits the GADD34/PP1 and CReP/PP1 complexes via direct binding or an indirect signaling event is currently under investigation.

The known chemical inhibitors of PP1 catalytic activity block the dephosphorylation of all PP1 substrates, including eIF2α (*23*). We therefore asked whether sal is selective in its inhibition of eIF2α dephosphorylation. Treatment of cells with a range of sal doses caused strong phosphorylation of eIF2α but not of Ser^7 on histone H3 (Fig. 3D). In contrast, treatment with calyculin A (CA), an active-site inhibitor of PP1 catalysis, caused parallel increases in both eIF2α and histone H3 phosphorylation (Fig. 3D). CA also induced Thr phosphorylation on many proteins in the cell, whereas sal did not induce Thr phosphorylation levels above vehicle control, even after long incubations (fig. S6). Furthermore, low doses of CA were toxic to PC12 cells (fig. S7), whereas doses of sal that induce similar levels of eIF2α phosphorylation (Fig. 3D) were nontoxic and cytoprotective (Fig. 1), confirming that the cellular effects of CA and sal are distinct. As a specificity control, the phosphorylation of the PP2A substrate extracellular signal–regulated kinase 2 (ERK2) (*24*) was unaffected by sal (fig. S8). Lastly, we used a proteomics approach (*25*) to look for global changes in protein level or posttranslational modification upon sal treatment. Of thousands of detectable proteins, only three to four were consistently affected by sal treatment (*26*). We concluded that sal selectively inhibits the PP1-mediated dephosphorylation of eIF2α and possibly a limited number of other substrates.

Protein phosphatases regulate many physiological and pathological processes. To test whether a selective pharmacological inhibitor of eIF2α dephosphorylation might be useful in a disease model, we turned to herpes simplex virus (HSV). Cellular PKR is activated by HSV infection and phosphorylates eIF2α to slow viral protein synthesis and virion production (*27*). To counteract this activity, HSV encodes ICP34.5, a protein homologous to GADD34, which binds cellular PP1 and mediates eIF2α dephosphorylation (*28*, *29*). Because ICP34.5 is essential for HSV replication in some cell types (*28*, *30*), we asked whether sal could inhibit ICP34.5/PP1-mediated eIF2α dephosphorylation and block viral replication. Sal induced eIF2α phosphorylation in both mock- and HSV-infected Vero cells (Fig. 4A), where ICP34.5/PP1 or PP1 is active (*29*). Consistent with previous reports (*29*), lysate of ICP34.5-transfected cells dephosphorylated eIF2α in vitro (Fig. 4B), whereas those of vector-transfected controls did not (fig. S4B). Sal treatment of ICP34.5-transfected cells inhibited dephosphorylation of eIF2α by the resulting lysate (Fig. 4B). We concluded that sal inhibits viral ICP34.5/PP1-mediated eIF2α dephosphorylation.

To determine whether sal inhibits HSV replication, we infected Vero cells with HSV in the presence of a dose range of sal and measured viral plaque formation. At high doses, sal reduced plaque formation by more than 95% [median inhibitory concentration (IC_{50}) ~ 3 μM] (Fig. 4C). To examine the role of eIF2α phosphorylation in the anti-

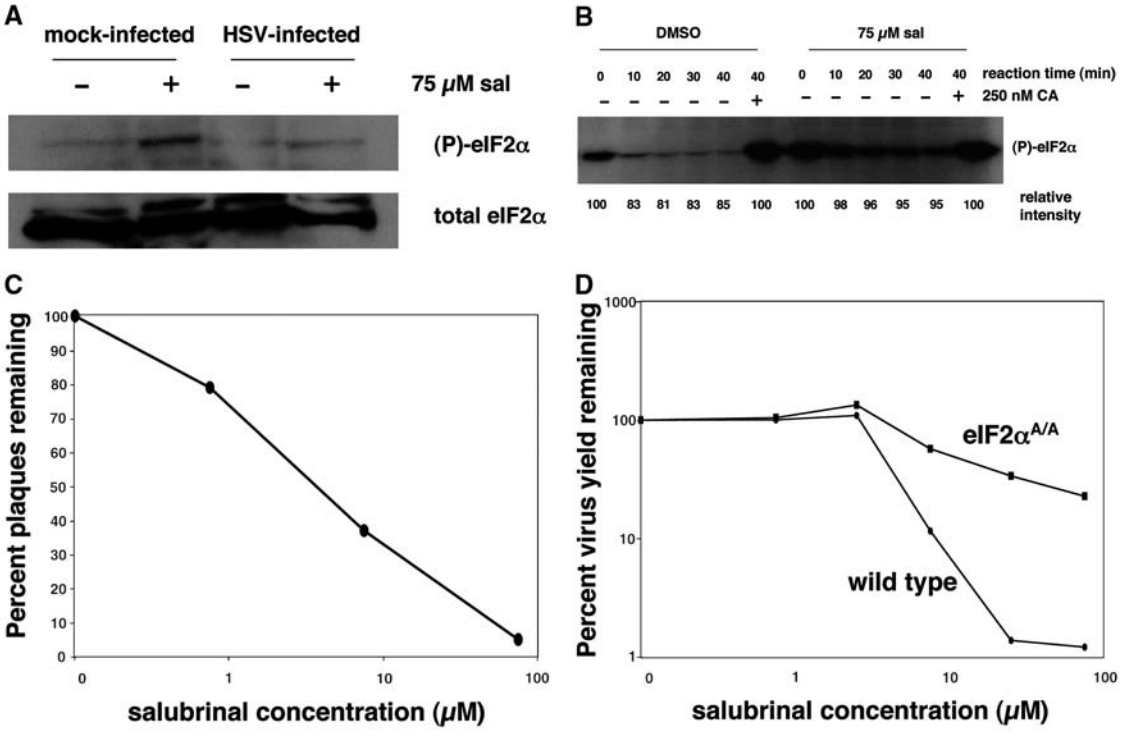


Fig. 4. Sal inhibits HSV replication by inhibiting eIF2α dephosphorylation. (**A**) Sal induced eIF2α phosphorylation in both mock-infected and HSV-infected Vero cells. Vero cells were treated with sal for 24 hours, infected with HSV, and harvested 18 hours postinfection (hpi). Protein levels were assessed by immunoblot. (**B**) Cell lysates made from ICP34.5-transfected CHO cells and treated with DMSO for 24 hours efficiently dephosphorylated ^{32}P-eIF2α labeled in vitro with γ-^{32}P-ATP and GST-PERK (lanes 1 to 6). Lysates from ICP34.5-expressing cells incubated with sal for 24 hours were partially inhibited in their ability to dephosphorylate eIF2α (lanes 7 to 12). CA (250 nM) serves as positive controls for inhibition of dephosphorylation. Reactions were analyzed by SDS-PAGE, and phosphorylation levels were assessed by autoradiography. (**C**) Vero cells were pretreated with sal as indicated for 24 hours before infection with an average of 150 plaque-forming units HSV per well. After 2 to 3 days in the continued presence of sal, viral plaques were quantitated, showing that sal inhibited HSV replication in a dose-dependent manner. Wells treated with DMSO alone contained ~150 plaques. (**D**) Wild-type MEFs (circles) or eIF2α$^{A/A}$ MEFs (squares) were treated with the indicated concentrations of sal for 24 hours and then infected with HSV at multiplicity of infection equal to 10 in the continued presence of sal. After the virus adsorbed to the cells, medium containing the indicated concentrations of sal was added. The virus produced was titered on Vero cells.

HSV activity of sal, we used MEFs homozygous for a $Ser^{51} \rightarrow Ala^{51}$ mutation in eIF2α ($eIF2\alpha^{A/A}$), which eliminates its phosphorylation site (*15*). Sal strongly inhibited HSV replication in wild-type MEFs, but inhibition was greatly reduced in $eIF2\alpha^{A/A}$ MEFs (Fig. 4D), demonstrating that induction of eIF2α phosphorylation is important for sal's anti-HSV effect. The residual inhibition of HSV replication in $eIF2\alpha^{A/A}$ MEFs may be due to sal's effect on as-yet unidentified substrates of ICP34.5/PP1 besides eIF2α. Lastly, we asked whether sal could inhibit HSV replication in a mouse cornea infection model (*31*). Compared to vehicle control, topical sal treatment significantly reduced the viral titer recovered from eye swabs of infected animals ($n = 8$, $P < 0.008$) (fig. S9). We concluded that sal can inhibit HSV replication in both cultured cells and an animal model of viral infection.

Because the catalytic subunits of most protein phosphatases show little intrinsic substrate preference, the selective pharmacological inhibition of protein dephosphorylations has generally been regarded as difficult or impossible. We have identified a small molecule with low toxicity that selectively inhibits a known PP1-mediated dephosphorylation in intact cells and used it to manipulate cellular physiology to achieve specific outcomes. Therefore, our results provide a proof-of-principle demonstration that pharmacological intervention against the dephosphorylation of selected substrates is a feasible approach for the development of research reagents and perhaps therapeutic agents. We show that selective small molecule inhibition of eIF2α dephosphorylation effectively protects cells from ER stress, revealing a critical role for eIF2α dephosphorylation in ER stress–induced apoptotic signaling, and blocks the replication of HSV, a widespread human pathogen. Future drugs that target eIF2α dephosphorylation may therefore complement the anti-HSV agents in current clinical use. More broadly, a detailed understanding of sal's mechanism may facilitate the discovery of inhibitors of other selected cell- or pathogen-directed dephosphorylations. Such small molecules could find therapeutic use in diseases as diverse as neurodegeneration, cancer, and viral infection.

References and Notes

1. R. J. Kaufman, *Genes Dev.* **13**, 1211 (1999).
2. H. P. Harding, M. Calfon, F. Urano, I. Novoa, D. Ron, *Annu. Rev. Cell Dev. Biol.* **18**, 575 (2002).
3. M. Aridor, W. E. Balch, *Nat. Med.* **5**, 745 (1999).
4. R. V. Rao, H. M. Ellerby, D. E. Bredesen, *Cell Death Differ.* **11**, 372 (2004).
5. R. V. Rao *et al.*, *J. Biol. Chem.* **276**, 33869 (2001).
6. J. A. Morris, A. J. Dorner, C. A. Edwards, L. M. Hendershot, R. J. Kaufman, *J. Biol. Chem.* **272**, 4327 (1997).
7. H. Yoshida, T. Matsui, A. Yamamoto, T. Okada, K. Mori, *Cell* **107**, 881 (2001).
8. M. Calfon *et al.*, *Nature* **415**, 92 (2002).
9. H. P. Harding, Y. Zhang, D. Ron, *Nature* **397**, 271 (1999).
10. R. Sood, A. C. Porter, D. A. Olsen, D. R. Cavener, R. C. Wek, *Genetics* **154**, 787 (2000).
11. H. P. Harding *et al.*, *Mol. Cell* **6**, 1099 (2000).
12. J. W. Brewer, J. A. Diehl, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 12625 (2000).
13. I. Novoa, H. Zeng, H. P. Harding, D. Ron, *J. Cell Biol.* **153**, 1011 (2001).
14. H. P. Harding, Y. Zhang, A. Bertolotti, H. Zeng, D. Ron, *Mol. Cell* **5**, 897 (2000).
15. D. Scheuner *et al.*, *Mol. Cell* **7**, 1165 (2001).
16. C. Jousse *et al.*, *J. Cell Biol.* **163**, 767 (2003).
17. P. D. Lu *et al.*, *EMBO J.* **23**, 169 (2004).
18. A. Hinnebusch, in *Translational Control of Gene Expression*, N. Sonenberg, J. W. B. Hershey, M. B. Mathews, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2000), pp. 185–244.
19. R. J. Kaufman, in *Translational Control of Gene Expression*, N. Sonenberg, J. W. B. Hershey, M. B. Mathews, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2000), pp. 503–528.
20. J. J. Chen, in *Translational Control of Gene Expression*, N. Sonenberg, J. W. B. Hershey, M. B. Mathews, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2000), pp. 529–546.
21. M. Boyce *et al.*, unpublished data.
22. X. Z. Wang *et al.*, *EMBO J.* **17**, 5708 (1998).
23. A. McCluskey, A. T. Sim, J. A. Sakoff, *J. Med. Chem.* **45**, 1151 (2002).
24. N. G. Anderson, J. L. Maller, N. K. Tonks, T. W. Sturgill, *Nature* **343**, 651 (1990).
25. M. Unlu, M. E. Morgan, J. S. Minden, *Electrophoresis* **18**, 2071 (1997).
26. M. Boyce, J. S. Minden, J. Yuan, unpublished observations.
27. J. Chou, J. J. Chen, M. Gross, B. Roizman, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10516 (1995).
28. J. Chou, B. Roizman, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 5247 (1994).
29. B. He, M. Gross, B. Roizman, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 843 (1997).
30. J. Chou, E. R. Kern, R. J. Whitley, B. Roizman, *Science* **250**, 1262 (1990).
31. D. A. Leib *et al.*, *J. Virol.* **63**, 759 (1989).

32. We thank J. Alvarez, J.-J. Chen, J. Glaven, O. Gozani, R. King, T. Mitchison, F. Roth, S. Ryeom, M. Shair, C.T. Walsh, and members of the Yuan lab for reagents and advice and J. Follen and D. Hayes of the Harvard Institute of Chemistry and Cell Biology for tireless help with small molecule screening. Supported by NIH grants R37-AG012859 and GM64703 (J.Y.); DDK42394 (R.J.K.); AI26077, AI19838, and NS35138 (D.M.C.); and ES08681 and DK47119 (D.R.); NSF (M.B.); grant KGCX2-SW-209 from the Chinese Academy of Sciences; and grant 20228205 from the National Natural Science Foundation of China (D.M.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/307/5711/935/DC1
Materials and Methods
Figs. S1 to S10

23 June 2004; accepted 21 December 2004
10.1126/science.1101902